

## Loop Extrusion Mediates Physiological Locus Contraction for V(D)J Recombination

Hai-Qiang Dai\*<sup>#1</sup>, Hongli Hu\*<sup>1</sup>, Jiangman Lou<sup>1</sup>, Adam Yongxin Ye<sup>1</sup>, Aimee M. Chapdelaine-Williams<sup>1</sup>, Nia Kyritsis<sup>1</sup>, Yiwen Zhang<sup>1</sup>, Nicole Manfredonia<sup>1</sup>, Rachael Judson<sup>1</sup>, Huan Chen<sup>1</sup>, Kerstin Johnson<sup>1</sup>, Sherry Lin<sup>1</sup>, Zhaoqing Ba<sup>1</sup>, Andrea Conte<sup>3</sup>, Rafael Casellas<sup>3</sup>, Cheng-Sheng Lee<sup>#2</sup>, and Frederick W. Alt<sup>#1</sup>

<sup>1</sup>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.

<sup>2</sup>Institute of Molecular and Cellular Biology, Department of Life Sciences, Hsinchu, Taiwan, 30013, R.O.C.

<sup>3</sup>Lymphocyte Nuclear Biology, NIAMS, NIH, and Center of Cancer Research, NCI, NIH, Bethesda, MD 20892, USA.

\*These authors contributed equally

#Correspondence

alt@enders.tch.harvard.edu

cslee@life.nthu.edu.tw

hai-qiang.dai@childrens.harvard.edu

**Abstract (207 words) (total text: 2940 words)**

**Immunoglobulin heavy chain locus (*Igh*)  $V_H$ , D, and  $J_H$  gene segments are developmentally assembled into V(D)J exons. RAG endonuclease initiates V(D)J recombination by binding a  $J_H$ -recombination signal sequence (RSS) within a chromatin-based recombination center (RC) and then, in an orientation-dependent process, scans upstream D-containing chromatin presented by cohesin-mediated loop extrusion for convergent D-RSSs to initiate  $DJ_H$ -RC formation<sup>1,2</sup>. In primary pro-B cells, 100s of upstream  $V_H$ -associated RSSs, embedded in convergent orientation to the  $DJ_H$ -RC-RSS, gain proximity to the  $DJ_H$ -RC for  $V_H$ -to- $DJ_H$  joining via a mechanistically-undefined  $V_H$ -locus contraction process<sup>3-7</sup>. Here, we report that a 2.4 mega-base  $V_H$  locus inversion in primary pro-B cells nearly abrogates rearrangements of normally convergent  $V_H$ -RSSs and cryptic RSSs, even though locus contraction *per se* is maintained. Moreover, this inversion activated rearrangement of both cryptic  $V_H$ -locus RSSs normally in the opposite orientation and, unexpectedly, of normally-oriented cryptic RSSs within multiple, sequential upstream convergent-CBE domains. Primary pro-B cells had significantly reduced transcription of *Wapl*<sup>8</sup>, a cohesin-unloading factor, versus levels in *v-Ab1* pro-B lines that lack marked locus contraction or distal  $V_H$  rearrangements<sup>2,9-11</sup>. Correspondingly, *Wapl* depletion in *v-Ab1* lines activated  $V_H$ -locus contraction and orientation-specific RAG-scanning across the  $V_H$ -locus. Our findings indicate that locus contraction and physiological  $V_H$ -to- $DJ_H$  joining both are regulated via circumvention of CBE scanning impediments.**

The ability of dominant CBE contact loops over mega-base (Mb) distances genome-wide to be anchored predominantly by convergently-oriented CBEs versus CBEs in other orientations was proposed to be mediated by cohesin-mediated loop extrusion<sup>12,13</sup>, as opposed to potential diffusion-based mechanisms that cannot readily explain orientation-bias. Parallels between findings on RAG chromatin exploration during V(D)J recombination and loop extrusion models provided a foundation for the discovery of fundamental roles of loop extrusion in D to  $J_H$  recombination<sup>1,9,10,12-14</sup>. Indeed, while Ds have downstream and upstream RSSs in opposite orientations, only downstream D-RSSs convergently-oriented with initiating  $J_H$ -RSSs are used

for D to J<sub>H</sub> joining<sup>1</sup>. However, while all V<sub>H</sub>-RSSs are in convergent orientation with the upstream DJ<sub>H</sub>-RC-RSS<sup>9</sup>, potential access of the upstream V<sub>H</sub>S, at the D-to-J<sub>H</sub> rearrangement stage, via linear scanning is largely blocked by CBEs in intergenic control region 1 (IGCR1) between the V<sub>H</sub> and D loci<sup>9-11</sup>. In this regard, V<sub>H</sub> utilization across the 2.4 Mb V<sub>H</sub> locus has been associated with physical *Igh* locus contraction in pro-B cells, which is proposed to bring V<sub>H</sub>S into close orbit around the DJ<sub>H</sub>-RC to allow V<sub>H</sub>S to gain RC access via diffusion<sup>3-7</sup>. In its simplest form, such a diffusion-based model also predicts that V<sub>H</sub>-RSS orientation would not play a dominant role in V<sub>H</sub> utilization across the V<sub>H</sub> locus.

As an initial test of a diffusion-based versus linear scanning model for long-range V<sub>H</sub> utilization, we inverted the entire 2.4Mb V<sub>H</sub> locus upstream of the D-proximal V<sub>H</sub>81X in the genome of C57BL/6 mice, leaving V<sub>H</sub>81X in normal position and orientation as a control (Fig. 1a, Extended Data Fig. 1). In C57BL/6 mice, the entire V<sub>H</sub> locus has been sequenced, revealing 109 functional V<sub>H</sub>S with locations accurately mapped within four V<sub>H</sub> domains referred to from proximal to distal as: 7183/Q52, Middle, J558, and J558/3609, respectively<sup>15,16</sup> (Fig. 1a, top, Supplementary Table 1). This large scale inversion placed the distal J558/3609 V<sub>H</sub>S in the proximal V<sub>H</sub> location and the proximal 7183/Q52 V<sub>H</sub> locus, besides V<sub>H</sub>81X, in the distal location, with all 108 functional V<sub>H</sub>S and their RSSs across the locus now in inverted orientation relative to the DJ<sub>H</sub>-RC RSS (Fig. 1a middle, bottom). Based on our sensitive HTGTS-V(D)J-seq assay<sup>17,18</sup>, this inversion nearly abrogated rearrangements of the normally convergent V<sub>H</sub>-RSSs across the inverted region in primary bone marrow (BM) pro-B cells, with utilization of the normally oriented and positioned proximal V<sub>H</sub>81X actually doubling, perhaps due to lack of competition from immediately upstream V<sub>H</sub>S<sup>10</sup> (Fig. 1b, top versus bottom panel; Supplementary Table 1).

To examine effects of the inversion on interactions of V<sub>H</sub> locus sequences with the RC, we performed high-resolution 3C-HTGTS<sup>10</sup> with a RC bait in RAG1-deficient cultured BM pro-B cells with or without the inversion. RAG-deficient cells were used to prevent confounding V(D)J recombination events<sup>10</sup>. These studies revealed a series of robust peaks of RC interaction across the V<sub>H</sub> locus (Fig. 1c, labeled 1-15), a number of which, including those involving Pax5-activated intergenic repeat (PAIR) elements<sup>19</sup> (Fig. 1c, blue bars), are considered hallmarks of V<sub>H</sub> locus contraction<sup>5</sup>. Notably, all of the major RC-V<sub>H</sub> locus interaction peaks were maintained

within the inverted  $V_H$  locus, despite the huge positional exchanges between more proximal and more distal portions of the locus (Fig. 1c; labeled 1-15; Extended Data Fig. 2a-d). Indeed, the overall RC interactions within the different regions of the inverted locus appeared almost as a mirror-image of the patterns of the normally-oriented locus (Fig. 1c, upper and lower). Thus, the 2.4Mb inversion did not markedly disrupt  $V_H$ -locus contraction *per se*. In addition, cohesin binding patterns determined by ChIP-seq and germline  $V_H$  transcription patterns determined by GRO-seq across the different  $V_H$  domains were recapitulated in detail in the inverted  $V_H$ -locus, again largely as mirror images of those of the normal  $V_H$ -locus (Extended Data Fig. 3a-d; Supplementary Data 1). Also notable is that many of the major interaction peaks corresponded to high level peaks of sense/antisense transcription, although within the inverted  $V_H$  locus orientation of sense and antisense transcription relative to the RC is reversed (Extended Data Fig. 3d, bottom). However, despite maintenance of RC interactions, transcription patterns, and binding of key looping factors (Fig. 1c; Extended Data Fig. 2e and 3a-d; Supplementary Data 1), the inverted  $V_H$ -RSSs were barely used by the V(D)J recombination process. This finding is consistent with the notion that DJ<sub>H</sub>-RC-bound RAG linear scanning of  $V_H$  locus chromatin plays an important role in  $V_H$  utilization in  $V_H$ -locus contracted BM pro-B cells.

It remained conceivable that the large-scale inversion could have somehow prevented  $V_H$ s across the inverted locus from accessing the DJ<sub>H</sub>-RC via diffusion following locus contraction. Introduction of ectopic RCs into non-antigen receptor loci across the genome in *v-Abl* pro-B lines allows RAG, in the absence of other major scanning impediments, to linearly explore convergent CBE-anchored chromatin domains, up to at least several Mb distances, until scanning is terminated at convergent CBE loop anchors<sup>9</sup>. In these experiments, orientation of initiating *bona fide* RSSs programs RAG activity linearly in one direction or the other, up to Mb distances, with linear scanning evidenced by low-level of utilization 100s of cryptic RSSs, as short as the conserved first 3 bp of the RSSs (CAC), but only when they are in convergent orientation with the initiating *bona fide* RSS<sup>9</sup>. In such experiments, cryptic RSSs in the same orientation as the RC-RSS (e.g. GTGs) are not used, demonstrating that cryptic RSS access ectopic RCs by linear scanning versus diffusion<sup>9</sup>. Thus, to further explore the predictions of a linear RAG scanning model for  $V_H$  utilization in BM pro-B cells, we analyzed RAG utilization of



cryptic RSSs across the  $V_H$  locus in BM pro-B cells (Fig. 2a, top). These analyses clearly demonstrated that cryptic RSSs across the length of 2.4 Mb  $V_H$  locus in convergent orientation with the  $DJ_H$ -RC-RSS were utilized, but those in the same orientation as the  $DJ_H$ -RC-RSS were not, strongly indicating that RC-bound RAG linearly scans across the  $V_H$  locus in BM pro-B cells (Fig. 2b, top and 2c; Extended Data Fig. 4a, top, and 4b). This analysis also revealed that cryptic RSS utilization did not extend far beyond the *Igh* locus, likely due to scanning activity largely being terminated within the  $V_H$  locus in most BM pro-B cells by robust rearrangement of  $V_H$ s with convergent *bona fide* RSSs (Fig. 1b, top; Fig. 2b, top and 2c; Extended Data Fig. 4a, top, and 4b).

The large  $V_H$  inversion locus provided the opportunity for an even more stringent test of the linear scanning model. Thus, the scanning model predicts within the inverted  $V_H$  locus, cryptic RSSs that normally were not utilized because they were in the same orientation as the  $DJ_H$ -RC-RSS would now be more robustly utilized due to being placed in convergent orientation by the inversion (Fig. 2a, bottom). Indeed, the 2.4Mb  $V_H$ -locus inversion essentially eliminated utilization of the previously convergently-oriented cryptic RSSs now placed in the same orientation as the  $DJ_H$ -RC-RSSs and activated utilization of cryptic RSSs previously in the same orientation that were now placed in convergent orientation (Fig. 2b, bottom and 2c; Extended Data Fig. 4a, bottom, 4c). Strikingly, the inversion also resulted in RAG scanning well beyond the end of the *Igh* locus in BM pro-B cells over distances of 4 Mb upstream, nearly to the telomere (Fig. 2b, bottom and 2c; Extended Data Fig. 4a, bottom, 4c). Moreover, consistent with linear RAG scanning through the long non-inverted region upstream of the  $V_H$  locus, cryptic RSS utilization indeed involved those normally in convergent orientation with the  $DJ_H$ -RC-RSS (Fig. 2b, bottom and 2c; Extended Data Fig. 4a, bottom, 4c). This continuation of RAG scanning upstream of the  $V_H$  locus in the context of the  $V_H$  locus inversion suggests that substantial inactivation of dominant  $V_H$  rearrangements, due to their inversion, allows  $DJ_H$ -RC-initiated RAG scanning to more robustly proceed through the  $V_H$  locus into the upstream region. RAG scanning through the CBE-rich *Igh* locus in BM pro-B cells with normal or inverted *Igh* loci, as well as RAG scanning through multiple upstream convergent CBE-based domains, which are well-known to impede scanning<sup>9,20-22</sup>, suggests that BM pro-B cells may have an active

mechanism to neutralize ability of  $V_H$ -associated CBEs to impede RAG scanning that influences convergent CBE domains more broadly than those just within the  $V_H$  locus.

G1-arrested *v-Abl* pro-B cells can be viably arrested in the G1 cell-cycle stage by treatment with *Abl* tyrosine kinase inhibitor (STI-571)<sup>23</sup>, during which time they induce RAG expression and undergo robust D to  $J_H$  joining, but very little  $V_H$ -to-D $J_H$  joining in association with little  $V_H$ -locus contraction<sup>1,10</sup>. Thus, as one approach to elucidate a potential CBE-neutralizing mechanism in normal BM pro-B cells, we employed GRO-seq to measure the transcriptional activity of various genes encoding proteins implicated in cohesin complex function. We did this in locus-contracted RAG1-deficient C57BL/6 mouse BM pro-B cells and G1-arrested *v-Abl*-transformed *E $\mu$ -Bcl2*-expressing pro-B cells ("*v-Abl* pro-B cells"), which were also derived from RAG1-deficient C57BL/6 mouse BM pro-B cells (Extended Data Fig. 5a). Notably, of all factors analyzed, we found only the Wings apart-like (*Wapl*) factor<sup>8</sup> to have very low transcription in BM pro-B cells versus *v-Abl* pro-B cells (Fig. 3a, b). *Wapl* is known to unload cohesin from chromosomes by transiently opening the cohesin ring<sup>24-26</sup> and, thereby, contributes to regulation of chromatin structure and protects against chromosome segregation errors and aneuploidy<sup>27,28</sup>. Cohesin also is positioned in genomes by transcription, CTCF and *Wapl*<sup>29,30</sup>. Most relevantly, *Wapl*-depletion extends CTCF-anchored loop sizes in mammalian cells in association with increasing cohesin density<sup>31</sup>, which may increase the probability of extending loop extrusion past dynamic CTCF-bound CBE loop anchors<sup>32,33</sup>. Based on the finding of relatively low *Wapl* transcription in BM pro-B cells, we tested effects of substantial *Wapl*-depletion on *Igh* loop domain formation and RAG scanning in G1-arrested C57BL/6 mouse *v-Abl* pro-B cells.

We generated a RAG1-deficient C57BL/6 *v-Abl* pro-B cell line that expressed the TIR1 F-box protein from *Oryza sativa* (OsTIR1) under control of the conditional Tet promoter (Tet-on-OsTIR1)<sup>34</sup>. OsTIR1 binds a mini auxin-inducible degron (mAID) in the presence of auxin to trigger proteasome-dependent degradation<sup>34</sup>. We refer to these modified cells as the "primary" *v-Abl* line (Extended Data Fig. 5b, c). We subsequently targeted a stop codon into both *Wapl* alleles in the primary *v-Abl* line to introduce in-frame sequences encoding mAID with a fluorescent monomeric Clover (mClover) cassette<sup>34</sup> (Extended Data Fig. 5d, e), and refer to the

resulting lines as "Wapl-degron *v-Abl* pro-B cells". For experiments, we added doxycycline (Dox) and auxin (Indole-3-acetic acid, IAA) to induce OsTIR1 expression and trigger proteasome-dependent degradation of the Wapl-mAID-mClover protein, respectively. To specifically deplete mAID-tagged Wapl protein in the G1 stage, we treated the Wapl-degron *v-Abl* pro-B cells with IAA and Dox from day1 to day4 during the STI-571-induced four-day G1 arrest experimental period (day0-4) (Extended Data Fig. 5f). Western blotting revealed that mAID-tagged Wapl was substantially depleted in G1-arrested IAA/Dox-treated, but not untreated, Wapl-degron *v-Abl* pro-B cells (Extended Data Fig. 5g, h). While Wapl depletion is lethal to dividing cells<sup>27,28</sup>, Wapl-depleted *v-Abl* pro-B cells underwent normal STI-571-induced G1 arrest with only modest effects on viability over the remaining experimental period (Extended Data Fig. 5i, j). Substantial depletion of chromatin-bound mAID-tagged Wapl across the *Igh* locus and genome-wide in G1-arrested Wapl-depleted *v-Abl* cells was confirmed by ChIP-seq (Extended Data Fig. 6a, b). Robust sense and anti-sense transcription through the RC, genome-wide transcription patterns, and binding of several key looping factors across V<sub>H</sub> domains were maintained in G1-arrested, IAA/Dox-treated Wapl-degron *v-Abl* cells (Extended Data Fig. 6c-e).

3C-HTGTS-based interaction experiments of untreated versus IAA/Dox-treated G1-arrested RAG1-deficient Wapl-degron *v-Abl* pro-B cells revealed low-level peaks of RC interactions in all but the proximal 7183/Q52 domain V<sub>H</sub> locus in untreated cells that were substantially increased upon Wapl depletion (Fig. 3c, d middle and bottom). Indeed, the major interactions peaks across the J558/3609, J558, and Middle V<sub>H</sub> domains that were substantially increased in IAA/Dox-treated Wapl-degron *v-Abl* pro-B cells closely corresponded to the dominant peaks observed in the context of locus-contracted RAG1-deficient, cultured BM pro-B cells (Fig. 3d, upper, labeled 1-15), indicating that Wapl-depletion in *v-Abl* lines enhanced V<sub>H</sub>-locus contraction (Fig. 3d). This finding supports the notion that locus contraction may, at least in part, involve modulation of ability of cohesin-mediated loop extrusion to circumvent CTCF-bound CBE impediments. Baits specific for either a prominent V<sub>H</sub>J558/3609 interaction peak or a prominent V<sub>H</sub>J558 interaction peak locale showed increased interactions with the RC upon Wapl-depletion in the absence of increased interactions with other *Igh* locales (Extended Data Fig. 7a, b). If extrapolated to be representative RC interactions of all separate V<sub>H</sub> locus

interaction sites across the  $V_H$  domains in Wapl depleted *v-Abl* lines, these latter findings predict that a substantial fraction of individual cells in the population would have one or another sequence across their  $V_H$  locus interacting with the RC.

To test effects of Wapl-depletion on *Igh* V(D)J recombination, we introduced RAG1 into RAG1-deficient C57BL/6 Wapl-degrogen cells (Extended Data Fig. 8a). Similar to primary *v-Abl* cells (Extended Data Fig. 8b, c), untreated Wapl degrogen *v-Abl* pro-B cells activated robust D to  $J_H$  rearrangement with  $J_H$ -distal DFL16.1 having highest rearrangement frequency and  $J_H$ -proximal DQ52 intermediate frequency (Fig. 3e, Supplementary Table 2). As anticipated, untreated cells had very little  $V_H$  to  $DJ_H$  rearrangement, with the vast majority of that which did occur involving the  $V_H$ -CBE dependent rearrangement of the very most proximal  $V_{HS}^{1,10}$  (Fig. 3f, upper; Extended Data Fig. 6e, upper). Wapl depletion markedly reduced distal DFL16.1 rearrangement levels, while promoting greatly increased proximal DQ52 rearrangement (Fig. 3e; Supplementary Table 2), a pattern also observed in the context of inactivation of IGCR1 CBEs<sup>10,35</sup> (Extended Data Fig. 8b, e, f; Supplementary Table 3). Strikingly, Wapl depletion promoted an approximately four-fold overall increased  $V_H$  to  $DJ_H$  recombination, associated with significantly increased utilization of  $V_{HS}$  across the four  $V_H$  domains of the 2.4 Mb *Igh* locus (Fig. 3f, bottom; Supplementary Table 2). However, an exception was that utilization of the very most proximal  $V_{HS}$  was significantly decreased (Fig. 3f, bottom; Supplementary Table 2), which is consistent with diminished impediment activity of the most proximal  $V_H$ -CBEs<sup>10</sup> (Extended Data Fig. 6e, upper; Extended Data Fig. 8d, bottom). Together, the interaction and V(D)J recombination data are consistent with Wapl extending loop extrusion past CTCF-bound IGCR1 CBEs and proximal  $V_H$ -CBEs. To further assess this apparently enhanced long-range RAG scanning, we analyzed RAG utilization of cryptic RSSs across the  $V_H$  locus in Wapl-depleted *v-Abl* pro-B cells. Consistent with the findings in locus-contracted BM pro-B cells (Fig. 2b, upper), in Wapl depleted *v-Abl* pro-B cells, RAG utilized cryptic RSSs across the length of 2.4 Mb  $V_H$  locus that were in convergence with, but not in the same orientation as, the initiating  $DJ_H$ -RC-RSS for joining (Extended Data Fig. 9). Together, these studies demonstrate that Wapl depletion facilitates orientation-specific scanning of RC-bound RAG across the entire  $V_H$  locus.

We demonstrate that RAG accesses most  $V_H$ s across the 2.4 Mb  $V_H$  locus in BM Pro-B cells by a linear, orientation-dependent scanning mechanism. Our finding of RC-based RAG scanning through multiple convergent CBE domains upstream of an inverted  $V_H$  locus suggests that pro-B cells may activate  $V_H$ -locus scanning via a mechanism that allows loop extrusion to bypass, at least partially, CTCF-bound CBEs more generally. Very low *Wapl* transcription in BM pro-B cells led us to deplete *Wapl* in *v-Abl* cells, which markedly enhanced both  $V_H$  locus contraction and long-range  $V_H$  utilization, indicating that both processes involve cohesin-mediated loop extrusion in the context of diminished CBE impediments. In this regard, our detailed analyses of individual  $V_H$ s are in accord with recent CTCF-depletion studies in *v-Abl* lines<sup>2</sup> that show residual CBE activity and/or transcription may enhance utilization of many  $V_H$ s along the scanning path (Supplementary Data 2). While RAG has been shown to scan directly over Mb distances across genomic sequences that lack major loop extrusion impediments<sup>9</sup>, other mechanisms for RAG access of distal  $V_H$ s by linear, orientation-dependent scanning of the impediment rich  $V_H$  locus are conceivable. For example, loop extrusion-mediated locus contraction may occur frequently in BM pro-B cells when RAG is not bound to the RC. In such cells, partial neutralization of CBE impediments could allow loop extrusion past the RC to proceed for varying distances across the locus (Fig. 3, legend). Then, upon RAG-binding to the RC, linear scanning could be initiated at the point extrusion has reached along the locus in a given pro-B cell. In this way, loop extrusion mediated-locus contraction could augment ability of RAG scanning to access more distal  $V_H$ s. Finally, our *Wapl*-depletion findings show that, beyond modulating CTCF activity<sup>2</sup>,  $V_H$  locus CBE neutralization can occur by modulating cohesin ability to circumvent CTCF-bound CBEs.

## REFERENCES

- 1 Zhang, Y. *et al.* The fundamental role of chromatin loop extrusion in physiological V(D)J recombination. *Nature* **573**, 600-604, doi:10.1038/s41586-019-1547-y (2019).
- 2 Ba Z, L. J., Ye A, Dai H-Q, Dring E, Lin S, Jain S, Kyritsis N, Kieffer-Kwon K-R, Casellas R & Alt FW. CTCF Orchestrates Long-Range Cohesin-driven V(D)J Recombinational Scanning. *bioRxiv* 2020.01.01.891473; doi: <https://doi.org/10.1101/2020.01.01.891473>; *Nature*, *In press* (2020).

- 3 Bossen, C., Mansson, R. & Murre, C. Chromatin topology and the regulation of antigen receptor assembly. *Annual review of immunology* **30**, 337-356, doi:10.1146/annurev-immunol-020711-075003 (2012).
- 4 Lucas, J. S., Zhang, Y., Dudko, O. K. & Murre, C. 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. *Cell* **158**, 339-352, doi:10.1016/j.cell.2014.05.036 (2014).
- 5 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, doi:10.1016/j.immuni.2013.08.011 (2013).
- 6 Proudhon, C., Hao, B., Raviram, R., Chaumeil, J. & Skok, J. A. Long-Range Regulation of V(D)J Recombination. *Advances in immunology* **128**, 123-182, doi:10.1016/bs.ai.2015.07.003 (2015).
- 7 Ebert, A., Hill, L. & Busslinger, M. Spatial Regulation of V-(D)J Recombination at Antigen Receptor Loci. *Advances in immunology* **128**, 93-121, doi:10.1016/bs.ai.2015.07.006 (2015).
- 8 Kueng, S. *et al.* Wapl controls the dynamic association of cohesin with chromatin. *Cell* **127**, 955-967, doi:10.1016/j.cell.2006.09.040 (2006).
- 9 Hu, J. *et al.* Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. *Cell* **163**, 947-959, doi:10.1016/j.cell.2015.10.016 (2015).
- 10 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 e114, doi:10.1016/j.cell.2018.04.035 (2018).
- 11 Guo, C. *et al.* CTCF-binding elements mediate control of V(D)J recombination. *Nature* **477**, 424-430, doi:10.1038/nature10495 (2011).
- 12 Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell reports* **15**, 2038-2049, doi:10.1016/j.celrep.2016.04.085 (2016).
- 13 Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6456-6465, doi:10.1073/pnas.1518552112 (2015).
- 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. *Advances in immunology* **139**, 93-135, doi:10.1016/bs.ai.2018.07.001 (2018).
- 15 Choi, N. M. *et al.* Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies. *Journal of immunology* **191**, 2393-2402, doi:10.4049/jimmunol.1301279 (2013).
- 16 Bolland, D. J. *et al.* Two Mutually Exclusive Local Chromatin States Drive Efficient V(D)J Recombination. *Cell reports* **15**, 2475-2487, doi:10.1016/j.celrep.2016.05.020 (2016).
- 17 Lin, S. G. *et al.* Highly sensitive and unbiased approach for elucidating antibody repertoires. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 7846-7851, doi:10.1073/pnas.1608649113 (2016).
- 18 Hu, J. *et al.* Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nature protocols* **11**, 853-871, doi:10.1038/nprot.2016.043 (2016).



- 19 Ebert, A. *et al.* The distal V(H) gene cluster of the Igh locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells. *Immunity* **34**, 175-187, doi:10.1016/j.immuni.2011.02.005 (2011).
- 20 Montefiori, L. *et al.* Extremely Long-Range Chromatin Loops Link Topological Domains to Facilitate a Diverse Antibody Repertoire. *Cell reports* **14**, 896-906, doi:10.1016/j.celrep.2015.12.083 (2016).
- 21 Lin, Y. C. *et al.* Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nature immunology* **13**, 1196-1204, doi:10.1038/ni.2432 (2012).
- 22 Benner, C., Isoda, T. & Murre, C. New roles for DNA cytosine modification, eRNA, anchors, and superanchors in developing B cell progenitors. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 12776-12781, doi:10.1073/pnas.1512995112 (2015).
- 23 Bredemeyer, A. L. *et al.* ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* **442**, 466-470, doi:10.1038/nature04866 (2006).
- 24 Buheitel, J. & Stemmann, O. Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3-Scc1 gate. *The EMBO journal* **32**, 666-676, doi:10.1038/emboj.2013.7 (2013).
- 25 Eichinger, C. S., Kurze, A., Oliveira, R. A. & Nasmyth, K. Disengaging the Smc3/kleisin interface releases cohesin from Drosophila chromosomes during interphase and mitosis. *The EMBO journal* **32**, 656-665, doi:10.1038/emboj.2012.346 (2013).
- 26 Chan, K. L. *et al.* Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell* **150**, 961-974, doi:10.1016/j.cell.2012.07.028 (2012).
- 27 Tedeschi, A. *et al.* Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature* **501**, 564-568, doi:10.1038/nature12471 (2013).
- 28 Haarhuis, J. H. *et al.* WAPL-mediated removal of cohesin protects against segregation errors and aneuploidy. *Current biology : CB* **23**, 2071-2077, doi:10.1016/j.cub.2013.09.003 (2013).
- 29 Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503-507, doi:10.1038/nature22063 (2017).
- 30 Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *The EMBO journal* **36**, 3573-3599, doi:10.15252/embj.201798004 (2017).
- 31 Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693-707 e614, doi:10.1016/j.cell.2017.04.013 (2017).
- 32 Wutz, G. *et al.* ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin(STAG1) from WAPL. *eLife* **9**, doi:10.7554/eLife.52091 (2020).
- 33 Li, Y. *et al.* The structural basis for cohesin-CTCF-anchored loops. *Nature* **578**, 472-476, doi:10.1038/s41586-019-1910-z (2020).
- 34 Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M. T. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell reports* **15**, 210-218, doi:10.1016/j.celrep.2016.03.001 (2016).

- 35 Qiu, X. *et al.* Sequential Enhancer Sequestration Dysregulates Recombination Center Formation at the IgH Locus. *Molecular cell* **70**, 21-33 e26, doi:10.1016/j.molcel.2018.02.020 (2018).

## ACKNOWLEDGEMENTS

We thank lab members for stimulating discussions, Hwei-Ling Cheng for advice and help with ES cell culture, Ming Tian for EF1 ES cell line, Yu Zhang and Xuefei Zhang for RAG-expressing retrovirus plasmids. This work was supported by NIH R01 AI020047 (to F.W.A.). H.-Q.D. is a fellow of the Cancer Research Institute (CRI) of New York. H.C. is a NRSA Fellow (T32 ai07386) and was supported by a Leukemia and Lymphoma Society fellowship. Z.B. was supported by CRI fellowship. R.C. and A.C. are partially funded by the NIH Regulome Project. C.-S.L. was previously supported by CRI fellowship and is now funded by the Ministry of Science and Technology in Taiwan [MOST109-2636-B-007-004]. F.W.A. is an investigator of the Howard Hughes Medical Institute.

## AUTHOR CONTRIBUTIONS

H.-Q.D., C.-S.L. and F.W.A. conceived the original idea and designed the study. H.-Q.D. generated and analysed the *Igh* V<sub>H</sub> locus inversion mouse model, and found the significantly low-level of *Wapl* transcription in BM pro-B cells. C.-S.L. initiated and established the *Wapl*-degron *v-Abl* system and obtained preliminary V usage and interaction results, and together with H.-Q.D. generated the doxycycline inducible *Wapl*-degron lines. H.-Q.D. optimized the RAG complementation method in the *Wapl*-degron system and performed the analyses, and together with H.L. and C.-S.L. performed 3C-HTGTS. H.L. performed ChIP-seq, Western blot and Southern blot. J.L. performed GRO-seq. A.Y.Y. and N.K. designed some of the bioinformatics pipelines for data analyses. A.M.C.-W with the assistance of N.M., R.J. and K.J. performed RDBC injections. H.C. optimized HTGTS-V(D)J-seq. Y.Z. assisted with the analysis of data. Z.B. and S.L. established the RAG1-deficient, *Eμ-Bcl2*-expressing C57BL/6 *v-Abl* pro-B cell line. H.-Q.D., H.L. and F.W.A., with input from C.-S.L., A.Y.Y., analyzed and interpreted data, designed figures, and wrote the manuscript. C.-S.L., R.C., J.L., Y.Z., Z.B. and A.C. provided insights and helped polish the paper. F.W.A. supervised the study.

## AUTHOR INFORMATION

The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.W.A. (alt@enders.tch.harvard.edu). F.W.A. is a co-founder of Otoro Biopharmaceuticals.



## FIGURE LEGENDS

### Figure 1 | *Igh* V<sub>H</sub> inversion mutation marked decreased utilization of all V<sub>H</sub>s other than the still convergently oriented V<sub>H</sub>81X in primary BM pro-B cells.

**a**, Schematic of the strategy to generate the mouse model with entire *Igh* V<sub>H</sub> locus inversion upstream V<sub>H</sub>81X. Diagram of the murine *Igh* locus showing the first two proximal V<sub>H</sub>S (V<sub>H</sub>81X and V<sub>H</sub>2-2), the last two distal V<sub>H</sub>S (V<sub>H</sub>1-86P and V<sub>H</sub>1-85), D<sub>S</sub>, J<sub>H</sub>S, C<sub>H</sub>S, and regulatory elements as indicated (not to scale), with RC that comprises J<sub>H</sub>-proximal DQ52 segment, four J<sub>H</sub> segments, and the intronic enhancer (iE<sub>μ</sub>) highlighted. All V<sub>H</sub> segments divided into four V<sub>H</sub>S domains (7183/Q52, Middle, J558 and J558/3609) from most proximal to distal. Yellow and dark orange triangles represent position and orientation of *bona fide* 23RSS and 12RSS, respectively. Purple and pink trapezoids represent position and orientation of CBEs. Green arrows denote the J<sub>H</sub>1-4 coding end bait primers used for generating HTGTS-V(D)J-seq libraries. Cut1 and Cut2 showing the location of 2 sgRNAs. Tel, telomere. Cen, centromere. **b**, Average utilization frequencies ± s.d. of all V<sub>H</sub> segments in WT (top, blue, n=3) and *Igh* V<sub>H</sub> inversion (bottom, red, n=4) mouse BM pro-B cells. Average percentage ± s.d. of V<sub>H</sub>DJ<sub>H</sub> and DJ<sub>H</sub> rearrangements are indicated. For comparison, several highly utilized V<sub>H</sub>S in each V<sub>H</sub> domain are indicated. See Supplementary Table 1 and Methods for more details. **c**, Average 3C-HTGTS signal counts ± s.e.m. across the four V<sub>H</sub> domains in cultured WT (top, n=3) and *Igh* V<sub>H</sub> inversion (bottom, n=3) *Rag1*<sup>-/-</sup> mouse BM pro-B cells. The WT and inverted V<sub>H</sub> locus/domains with PAIR elements are diagrammed at the top of each panel. For comparison, the location of the highly utilized V<sub>H</sub>S in (b) are labeled below each panel (c). 15 representative major interaction peaks/clusters were indicated with colour shades and numbers. See Extended Data Fig. 2 and Supplementary Data 1 for more details. In all panels, “n” indicates biological repeats. *p* values were calculated using unpaired two-tailed t-test.

### Figure 2 | RAG targets cryptic RSSs that *Igh* locus inversion place in convergent orientation to the upstream D-RSSs in primary BM pro-B cells.

**a**, Illustration of possible joining outcomes between *bona fide* RSSs from J<sub>H</sub>1-4 coding end baits to cryptic RSSs mostly represented by CAC motifs in the upstream D<sub>S</sub>, V<sub>H</sub>S and upstream region till telomere in WT (top) and *Igh* V<sub>H</sub> inversion (bottom) mouse primary pro-B cells. Red and blue arches with arrows show possible deletional and inversional junctions, respectively, except for the *Igh* V<sub>H</sub> inverted region (red for inversional, blue for deletional). Green arrows indicate the position and orientation of HTGTS primers. Details as shown in Fig. 1a. **b**, RAG off-target junction profiles at *Igh* locus and upstream 4 Mb region in WT (middle) and *Igh* V<sub>H</sub> inversion (bottom) BM pro-B cells isolated from individual mice. Top panel is WT 2.4 Mb *Igh*

locus and upstream 4 Mb region track. Junctions are displayed with linear scale as stacked tracks. I-VI, indicated by sky blue lines, are examples of RAG off-target peaks plotted at single-base-pair resolution in Extended Data Fig. 4a. The background of *Igh* V<sub>H</sub> inverted region is highlighted in transparent green (**a, b**, WT) and transparent gold (**a, b**, *Igh* V<sub>H</sub> inversion), respectively. Del (+), deletional junction. Inv (-), inversional junction. Each library was normalized. See methods and related Extended Data Fig. 4 for more details and additional independent repeats. **c**, Average frequencies  $\pm$  s.d. of plus strand (red, +) and minus strand (blue, -) joining events within indicated regions in wide-type (WT,  $n=3$ ) and *Igh* V<sub>H</sub> inversion (INV,  $n=3$ ) mouse BM pro-B cells. Three indicated regions (J<sub>H</sub>4-V<sub>H</sub>81X locus, Entire *Igh* upstream V<sub>H</sub>81X locus, and Entire upstream *Igh* locus) were highlighted (**b**), respectively. Each of the biological library replicates was normalized to 2979 off-target junctions for statistical analysis. *p* values were calculated using unpaired two-tailed t-test. See Methods for more details.

**Figure 3 | Wapl depletion activates *Igh* RC interactions with the V<sub>H</sub> locus and results in increased distal V<sub>H</sub>S utilization in G1-arrested Wapl-degrogen *v-Abl* pro-B cells.**

**a**, Scatter plots of averaging transcriptome-wide GRO-seq counts in G1-arrested *v-Abl* pro-B cells (*x axis*,  $n=3$ ) and BM pro-B cells (*y axis*,  $n=4$ ). Representative known requisite genes implicated in the cohesin complex function for V(D)J recombination and chromatin interaction are highlighted by red circles and blue arrows. Spearman's correlation coefficient ( $\rho$ ) and *p* value determined by two-sided Spearman's correlation test are presented. **b**, Comparison of *Wapl* transcription level in BM pro-B cells and G1-arrested *v-Abl* pro-B cells from 4 and 3 independent repeats, respectively. The transcription level of genes lying upstream and downstream of *Wapl* were shown in the same panel. **c**, Diagram of the entire murine *Igh* locus with details as shown in Fig. 1a. Purple and green arrows denote the RC/*iE $\mu$*  bait and J<sub>H</sub>1-4 coding end bait primers used for 3C-HTGTS and HTGTS-V(D)J-seq, respectively. **d**, Average 3C-HTGTS signal counts  $\pm$  s.e.m. across the four V<sub>H</sub> domains in cultured *Rag1*<sup>-/-</sup> BM pro-B cells (top,  $n=3$ ), G1-arrested *Rag1*<sup>-/-</sup> *v-Abl* pro-B cells without (middle,  $n=3$ ) or with *Wapl* depletion (bottom,  $n=3$ ). For comparison, 15 representative major interaction peaks/clusters as shown in Fig.1c were indicated with colour shades and numbers. PAIR elements are indicated as blue bars. **e, f**, Average utilization frequencies  $\pm$  s.d. of all D segments (**e**) from DJ<sub>H</sub> and V<sub>H</sub>DJ<sub>H</sub> joins, and all V<sub>H</sub> segments (**f**) in RAG1-complemented, G1-arrested *Wapl*-degrogen *v-Abl* pro-B cells without (Untreated,  $n=3$ ) or with (IAA&Dox, red,  $n=3$ ) *Wapl* depletion. Average percentage  $\pm$  s.d. of V<sub>H</sub>DJ<sub>H</sub> and DJ<sub>H</sub> rearrangements are indicated (**f**). For comparison, several highly utilized V<sub>H</sub>S (**f**) in *Wapl*-depleted *v-Abl* pro-B cells and the correlated interaction peaks (**d**) are highlighted,

respectively. *p* values were calculated using unpaired two-tailed t-test. See Supplementary Table 2 and Methods for more details. Notably, while Wapl-depleted *v-Ab1* cells have RC interactions across the  $V_H$  locus that are very similar to those of BM pro-B cells (**d**, upper and bottom); their  $V_H$  to  $DJ_H$  rearrangement levels are much lower than those of BM pro-B cells (compare Fig. 1b and 3f), despite having activated RAG scanning across the entire  $V_H$  locus (**d**). These observations may at least in part reflect RC-interactions being assayed in RAG-deficient cells while V(D)J recombination is assayed in RAG sufficient cells. In this regard, prior studies suggested extrusion of chromatin impediments past the RC may be more efficient when it is not RAG-bound<sup>10</sup>.

## EXTENDED DATA FIGURE LEGENDS

### Extended Data Figure 1 | Generation and characterization of *Igh* $V_H$ locus inversion mouse model.

**a**, Schematic diagram showing CRISPR-Cas9-mediated entire *Igh*  $V_H$  locus inversion upstream  $V_H81X$  in embryonic stem cells (ES cells) on the *Igh* allele in C57BL/6 genetic background. Cut1 and Cut2 showing the location of 2 sgRNAs. Details as shown in Fig. 1a. **b**, Confirmation of the upstream and downstream inversion junctions by Sanger sequencing. The sgRNA-targeting sequence is underlined, and the PAM sequence is labelled in red. sgRNAs and oligos used are listed in Supplementary Table 4. **c**, Schematic showing the generation of *Igh*  $V_H$  locus inversion mouse model and further assays for phenotype and mechanism analysis. **d, e**, Representative flow cytometry analysis of IgM<sup>-</sup> bone marrow (BM) B cell populations in 4~6-week-old WT (**d**) and *Igh*  $V_H$  locus inversion (**e**) mice. B220<sup>+</sup>IgM<sup>-</sup> B cells were gated and shown in the left plot (**d, e**). B220<sup>+</sup>CD43<sup>+</sup> pro-B and B220<sup>+</sup>CD43<sup>-</sup> pre-B cell populations are indicated in the right plot (**d, e**).

### Extended Data Figure 2 | 3C-HTGTS chromatin interaction profiles at *Igh* locus and upstream region in WT and *Igh* $V_H$ locus inversion cultured *Rag1*<sup>-/-</sup> BM pro-B cells.

For comparison, all the data were shown in normal *Igh* orientation. **a**, 3C-HTGTS chromatin interaction profiles of RC/iE $\mu$  bait across the 4  $V_H$  domains and 1.6 Mb upstream region in WT (blue, *n*=3) and *Igh*  $V_H$  locus inversion (red, *n*=3) cultured *Rag1*<sup>-/-</sup> mouse BM pro-B cells. **b-d**, 3C-HTGTS profiles zoomed in the upstream *Igh* region (**b**), distal *Igh*  $V_H$  region (**c**), and proximal *Igh*  $V_H$  region (**d**). **e**, 3C-HTGTS chromatin interaction profiles of distal  $V_H8-7P$  bait across the entire *Igh* locus in WT (blue, *n*=2) and *Igh*  $V_H$  locus inversion (red, *n*=2) cultured *Rag1*<sup>-/-</sup> mouse BM pro-B cells. All  $V_H$  segments divided into four  $V_{HS}$  domains (7183/Q52, Middle, J558 and J558/3609) from most proximal to distal. The background of *Igh*  $V_H$  inverted

region is highlighted in transparent green (WT) and transparent gold (*Igh* V<sub>H</sub> inversion), respectively. In all panels, “n” indicates biological repeats. See Methods for more details.

**Extended Data Figure 3 | Rad21-binding and transcription across *Igh* locus in WT and *Igh* V<sub>H</sub> locus inversion cultured *Rag1*<sup>-/-</sup> BM pro-B cells.**

**a**, Representative Rad21 ChIP-seq profiles across the 4 V<sub>H</sub> domains as indicated in WT (upper, blue) and *Igh* V<sub>H</sub> locus inversion (bottom, red) primary pro-B cells. **b, c**, Rad21 ChIP-seq profiles zoomed in the distal *Igh* V<sub>H</sub> region (**b**) and proximal *Igh* V<sub>H</sub> region (**c**). **d**, Average signal counts ± s.e.m. of GRO-seq profiles across the 4 V<sub>H</sub> domains in WT (upper) and *Igh* V<sub>H</sub> locus inversion (bottom) BM pro-B cells. The WT and inverted V<sub>H</sub> locus/domains with PAIR elements are diagrammed at the top of each panel. Both the sense and antisense transcription are relative to the entire *Igh* V<sub>H</sub> locus upstream V<sub>H</sub>81X with or without inversion and indicated respectively. For comparison, 15 representative major interaction peaks/clusters as shown in Fig.1c were indicated with colour shades and numbers. The background of *Igh* V<sub>H</sub> inverted region is highlighted in transparent green (WT) and transparent gold (*Igh* V<sub>H</sub> inversion), respectively. In all panels, “n” indicates biological repeats. See Methods for more details.

**Extended Data Figure 4 | RAG off-target profiles in WT and *Igh* V<sub>H</sub> locus inversion primary BM pro-B cells.**

**a**, 6 examples (labeled in Fig. 2b, I-VI) of RAG off-target peaks in WT (repeat#1, middle) and *Igh* V<sub>H</sub> locus inversion (repeat#1, bottom) BM pro-B cells plotted at single-base-pair resolution. Top panels are WT 2.4 Mb *Igh* locus and upstream 4 Mb region track. **b, c**, RAG off-target junction profiles at *Igh* locus and upstream 4 Mb region in WT (repeat#2-3, middle) and *Igh* V<sub>H</sub> locus inversion (repeat#2-3, bottom) BM pro-B cells isolated from individual mice. The same regions of RAG off-target peaks (shown in **a**) are highlighted with sky blue lines. #1-3, number of individual WT or *Igh* V<sub>H</sub> inversion mice. The background of *Igh* V<sub>H</sub> inverted region is highlighted in transparent green (WT) and transparent gold (*Igh* V<sub>H</sub> inversion), respectively. Del (+), deletional junction. Inv (-), inversional junction.

**Extended Data Figure 5 | Generation and characterization of Wapl-degrom *v-Abl* pro-B lines.**

**a**, Strategy for detecting differentially expressed genes between BM pro-B cells and G1-arrested *v-Abl* pro-B cells by GRO-seq. 24 interested genes are listed. **b**, Schematic of the targeting strategy for introducing Tet-On OsTIR1 expression cassette at the mouse *Rosa26* locus. Positions of homology arms (gray box) are indicated. Cells were transfected with pTet-On\_OsTIR1\_V5 and pX330\_*Rosa26* to target *Rosa26* locus. The clones were screened by genomic PCR. See Methods for more details. **c**, Immunoblotting to detect OsTIR1 expression

with the induction of Doxycycline (Dox). The indicated clones were grown in the presence of 2 mg/ml Dox for 24 hr before immunoblotting with the anti-V5 antibody. Primary #5 clone was picked for further Wapl-degron targeting.  $\beta$ -actin was a loading control. **d**, Strategy to generate Wapl-degron *v-Ab1* pro-B cell lines. Cells were transfected with pW290 and pX330\_Wapl-AID to introduce in-frame mAID-Clover sequences into the last codon of mouse both endogenous *Wapl* alleles. Positions of homology arms (gray box), Cas9/sgRNAs and Southern blot probes are indicated. **e**, Southern blot confirmation of two correctly targeted clones (#5-3 and #5-9) with Wapl-mAID on both alleles. **f**, Diagram of the experimental strategy to specifically deplete mAID-tagged Wapl protein in G1-arrested *v-Ab1* cells. **g**, Immunoblotting to detect Wapl and Wapl-mAID-Clover (Wapl-mAC) protein. The indicated clones (#5-3 and #5-9) were grown without or with Wapl depletion at indicated time points before immunoblotting. The specific western blot bands of WT Wapl and Wapl-mAC were labeled. OsTIR1 was detected by anti-V5 antibody. Primary #5 clone was used for the WT Wapl control and  $\beta$ -actin was a loading control. **h**, Representative flow-cytometry plots showing the percentage of Clover-positive Wapl-degron *v-Ab1* cells that are without (Untreated) or with (IAA&Dox) Wapl depletion at different time points (day 0, top; day 2, middle; day 4, bottom). Primary line #5 was processed as a Clover-negative control. **i**, Time-course cell viability assay for G1-arrested *v-Ab1* pro-B cells without or with Wapl depletion at different time points. Average percentage  $\pm$  s.d. of viable cells for each timepoint and for each condition was calculated 4 independent experiments. **j**, Representative flow-cytometry plots of propidium iodide (PI) stained G1-arrested *v-Ab1* pro-B cells without or with Wapl depletion at indicated time points. Percentages in the top-right corner represent the percentage of cells at G1, S and G2/M stage. Plasmids, sgRNAs and oligos used are listed in Supplementary Table 4. *p* values were calculated using unpaired two-tailed t-test. See Methods for more details.

### **Extended Data Figure 6 | Characterization of Wapl/Rad21/CTCF-binding and genome-wide gene transcription in G1-arrested Wapl-degron *v-Ab1* pro-B cells.**

**a**, Average signal counts  $\pm$  s.e.m. of Wapl ChIP-seq across the *Igh* locus are plotted as indicated in *Rag1*<sup>-/-</sup> G1-arrested *v-Ab1* pro-B cells without (Untreated, blue, n=3) or with (IAA&Dox, red, n=3) Wapl depletion and cultured *Rag1*<sup>-/-</sup> BM pro-B cells (green, n=3). **b**, Three independent repeats of Wapl ChIP-seq signal within  $\pm 1.0$  kb region across all peaks genome-wide called in G1-arrested *Rag1*<sup>-/-</sup> *v-Ab1* pro-B cells without (Untreated) or with (IAA&Dox) Wapl depletion. Top: Average enrichment. IAA&Dox treatment leads to a depletion of chromatin-bound Wapl genome-wide. See Methods for details. **c**, Average signal counts  $\pm$  s.e.m. of GRO-seq across the 4 V<sub>H</sub> domains (left) and RC region (right) are plotted as indicated in G1-arrested

*Rag1*<sup>-/-</sup> *v-Abl* pro-B cells without (Untreated, upper, n=3) or with (IAA&Dox, middle, n=3) Wapl depletion and cultured *Rag1*<sup>-/-</sup> BM pro-B cells (bottom, n=4). For comparison, 15 representative major interaction peaks/clusters as Fig.1c were indicated with colour shades and numbers. PAIR elements are indicated as blue Bars. **d**, Scatter plots of transcriptome-wide GRO-seq counts from three independent repeats in G1-arrested *Rag1*<sup>-/-</sup> *v-Abl* pro-B cells without (Untreated, x axis) and with (IAA&Dox, y axis) Wapl depletion. Representative known requisite genes for V(D)J recombination and chromatin interaction are highlighted by red circles and blue arrows in each of the three scatter plots. See Methods for more details. **e**, Average signal counts ± s.e.m. of Rad21 (upper) and CTCF (bottom) ChIP-seq across the four V<sub>H</sub> domains are plotted as indicated in G1-arrested *Rag1*<sup>-/-</sup> *v-Abl* pro-B cells without (Untreated, blue, n=3) or with (IAA&Dox, red, n=3) Wapl depletion. See Methods for more details.

**Extended Data Figure 7 | Wapl depletion results in increased chromatin interactions of the RC with distal V<sub>H</sub>s in G1-arrested Wapl-degroun *v-Abl* pro-B cells.**

**a**, 3C-HTGTS chromatin interaction profiles of distal V<sub>H</sub>1-47 bait across the entire *Igh* locus in G1-arrested *Rag1*<sup>-/-</sup> *v-Abl* pro-B cells without (Untreated, blue, n=2) or with Wapl depletion (IAA&Dox, red, n=2). **b**, 3C-HTGTS chromatin interaction profiles of distal V<sub>H</sub>8-7P bait across the entire *Igh* locus in WT *Rag1*<sup>-/-</sup> mouse BM pro-B cells (BM, green, 2 repeats) and G1-arrested *Rag1*<sup>-/-</sup> *v-Abl* pro-B cells without (Untreated, blue, n=2) or with Wapl depletion (IAA&Dox, red, n=2). Related very low-level peaks of RC interactions in untreated cell suggest low-level *Igh* locus contraction in untreated cells. “n” indicates experimental repeats. 2 biological repeats for BM pro-B cells. See Methods for more details.

**Extended Data Figure 8 | HTGTS-V(D)J-seq assay in RAG1-complemented, G1-arrested Wapl-degroun *v-Abl* pro-B cells.**

**a**, Diagram of the experimental strategy with retrovirus-mediated RAG complementation in Wapl-degroun *v-Abl* cells for HTGTS-V(D)J-seq assay. See Methods for more details. **b**, Average utilization frequencies ± s.d. of all D segments from DJ<sub>H</sub> and V<sub>H</sub>DJ<sub>H</sub> joins in RAG-complemented, G1-arrested untreated primary *v-Abl* cells (primary, green, n=3), untreated single *Igh* Wapl-degroun *v-Abl* pro-B cells (single *Igh*, purple, n=3) and untreated single *Igh* with IGCR1 KO Wapl-degroun *v-Abl* pro-B cells (IGCR1 KO, blue, n=1). **c**, Average utilization frequencies ± s.d. of all V<sub>H</sub> segments in RAG1-complemented, G1-arrested primary *v-Abl* pro-B cells (n=3). Average percentage ± s.d. of V<sub>H</sub>DJ<sub>H</sub> and DJ<sub>H</sub> rearrangements are indicated. **d**, Average 3C-HTGTS signal counts ± s.e.m. baiting from RC/iEμ across the locus between 3’CBE and several proximal V<sub>s</sub> in cultured *Rag1*<sup>-/-</sup> BM pro-B cells (top, green, n=3), G1-arrested *Rag1*<sup>-/-</sup> *v-Abl* pro-B cells without (middle, blue, n=3) or with Wapl depletion (bottom, red, n=3).



Related very low-level peaks of RC interactions with IGCR1 in IAA/Dox treated cells suggest IGCR1 was neutralized in Wapl depleted *v-Abl* cells. **e**, Average utilization frequencies  $\pm$  s.d. of all  $V_H$  segments in RAG1-complemented, G1-arrested untreated single *Igh* Wapl-degrogen *v-Abl* pro-B cells (n=3). Average percentage  $\pm$  s.d. of  $V_HDJ_H$  and  $DJ_H$  rearrangements are indicated. **f**, Representative utilization frequencies of all  $V_H$  segments in RAG1-complemented, G1-arrested untreated single *Igh* with IGCR1 KO Wapl-degrogen *v-Abl* pro-B cells (n=1). Representative percentage of  $V_HDJ_H$  and  $DJ_H$  rearrangements are indicated. All  $V_H$  segments divided into four domains from most proximal to distal. *p* values were calculated using unpaired two-tailed t-test. See Supplementary Table 2, 3 and Methods for more details.

### **Extended Data Figure 9 | Substantial Wapl depletion promotes robust RAG utilization of cryptic RSSs across the $V_H$ locus in G1-arrested Wapl-degrogen *v-Abl* pro-B cells.**

**a**, Illustration of possible joining outcomes between *bona fide* RSSs from  $J_H1-4$  coding end baits to cryptic RSSs mostly represented by CAC motifs across *Igh* locus, details as shown in Fig2a. **b**, Pooled HTGTS junction profiles across highlighted  $V_H$  domains for deletional and inversional joining in RAG1-complemented, G1-arrested *v-Abl* pro-B cells without (Untreated, upper, pooled n=3) or with Wapl depletion (IAA&Dox, bottom, pooled n=3). For presentation, all libraries from untreated or IAA&Dox treated cells were pooled and normalized. See Methods for more details. The vast majority of deletional joining events across the  $V_H$  locus are reproducible among replicates and involve, at a minimum the CAC of an RSS. Almost all of the handful of the very low-level inversional joining events interspersed across  $V_H$  locus are not reproducible among replicates consistent with being background events. **c**, Average frequencies  $\pm$  s.d. of deletional (left) and inversional (right) joining events within indicated regions in RAG1-complemented, G1-arrested *v-Abl* pro-B cells without (Untreated, n=3) or with Wapl depletion (IAA&Dox, n=3). Each replicate was normalized to 500 off-target junctions for statistical analysis. Three indicated regions (IgH  $V_H$  region, IGCR1- $V_H81X$  and DQ52-DFL16.1) were labeled (**b**), respectively. Del (+), deletional junction. Inv (-), inversional junction. *p* values were calculated using unpaired two-tailed t-test. See Methods for more details.

### **Additional Information for Supplementary Tables and Data**

**Supplementary Table 1** | Relative utilization of *Igh* V and D segments in WT and *Igh*  $V_H$  locus inversion primary BM pro-B cells (see text).

**Supplementary Table 2** | Relative utilization of *Igh* V and D segments in RAG1-complemented, G1-arrested, primary *v-Abl* pro-B cells and Wapl-degrogen *v-Abl* pro-B cells without or with Wapl depletion (see text).

**Supplementary Table 3** | Relative utilization of *Igh* V and D segments in IGCR1-intact or IGCR1-deleted RAG1-complemented G1-arrested untreated single *Igh* Wapl-degron *v-Abl* pro-B cells (see text).

**Supplementary Table 4** | Oligos and plasmids used in this study.

**Supplementary Data 1** | 3C-HTGTS, GRO-seq, CHIP-seq, CBE motif sites and PAIR elements for 15 representative peaks/clusters in  $V_H$  domains in cultured *Rag1*<sup>-/-</sup> BM pro-B cells (see text).

**Supplementary Data 2** |  $V_H$  usage, GRO-seq, 3C-HTGTS, CHIP-seq, CBE motifs sites and PAIR elements for all  $V_H$ s in G1-arrested Wapl-degron *v-Abl* pro-B cells (see text).

## 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.

### Mice

Wild-type (WT) C57BL/6 mice were purchased from Charles River Laboratories International. *Rag1*<sup>-/-</sup> mice (stock No.002216: B6.129S7-*Rag1*<sup>tm1Mom</sup>/J) in C57BL/6 background were purchased from the Jackson Laboratory and maintained in the lab. All animal experiments were performed under protocols approved by the Institutional Animal Care and Use Committee of Boston Children's Hospital.

### Generation and characterization of *Igh* $V_H$ locus inversion mouse model

The CRISPR/Cas9-mediated entire *Igh*  $V_H$  locus inversion upstream  $V_H81X$  modifications were carried out on the C57BL/6 *Igh* allele in EF1 ES cell line, which was derived from a F1 hybrid mouse (129/Sv:C57BL/6). The ES targeting was performed as previously described<sup>36</sup>. The cells were cultured at 37 °C with 5% CO<sub>2</sub> in DMEM medium supplemented with 15% fetal bovine serum, 20 mM HEPES, 1× MEM nonessential amino acids, 2 mM Glutamine, 100 units of Penicillin/Streptomycin, 100 μM β-mercaptoethanol, 500 units/ml Leukemia Inhibitory Factor (LIF). The ES cells were grown on a monolayer of mouse embryonic fibroblasts (MEF) that have been mitotically inactivated with γ-irradiation. sgRNAs targeting the upstream of  $V_H1-86P$  (Cut1) and upstream of  $V_H81-X$  (downstream of  $V_H2-2$ ) (Cut2) were cloned by annealing pairs of oligos into pX459 with a puromycin selection marker (Addgene, #62988) following the standard protocol. For transfection, Cas9/sgRNAs (Cut1 and Cut2, each 2.5 μg) were nucleofected via Lonza 4D Nucleofector into 2.5×10<sup>6</sup> EF1 cells. The electroporated ES cells were grown on a monolayer of MEF cells with puromycin resistance (Stemcell, #00325) for 1 day. Then the ES



cells were selected with 2 µg/ml puromycin (Gibco, #A1113802) for another day. After selection, the ES cells were cultured with normal medium for 5-6 days, colonies were picked into 24 well plates that have been coated with MEF. After 3 days, half of the colonies were frozen at -80°C and the other half were expanded for DNA isolation. The genomic DNA was analyzed with screening-PCR to identify clones in which the desired genetic modifications have taken place. After the primary screen, the upstream and downstream of modified region of the potential positive clones was amplified by confirming-PCR and sequenced. After the *Igh* V<sub>H</sub> inversion modifications were confirmed on the C57BL/6 *Igh* allele by single nucleotide polymorphisms (SNP), each positive clone was further validated with a normal karyotype by metaphase analysis. All the cell lines were confirmed to be mycoplasma free. The ES clone with the *Igh* V<sub>H</sub> inversion modifications was injected into *Rag2*<sup>-/-</sup> blastocysts to generate RDBC chimeras<sup>37</sup>. The chimeric mice were bred with C57BL/6 mice for germline transmission of the targeted mutations. Heterozygous *Igh* V<sub>H</sub> inversion mice were bred to homozygosity (*Igh* V<sub>H</sub> inversion mouse model). For further analysis, this mouse model was bred to *Rag1*<sup>-/-</sup> homozygosity in C57BL/6 background. Sequences of all sgRNAs and oligos used are listed in Supplementary Table 4.

### **Bone marrow pro-B cells purification**

For RAG on-target and off-target analysis, single cell suspensions were derived from bone marrows of 4-6-week-old C57BL/6 WT and *Igh* V<sub>H</sub> locus inversion mice and incubated in Red Blood Cell Lysing Buffer (Sigma-Aldrich, #R7757) to deplete the erythrocytes. B220<sup>+</sup>CD43<sup>hi</sup>IgM<sup>-</sup> pro-B cells were isolated by staining with anti-B220-APC (eBioscience, #17-0452-83), anti-CD43-PE (BD PharMingen, #553271), and anti-IgM-FITC (eBioscience, #11-5790-81) antibodies for 30 minutes at 4 °C and then purified by FACS, the purified primary pro-B cells were subjected to HTGTS-V(D)J-seq.

For 3C-HTGTS, ChIP-seq and GRO-seq, B220-positive WT and *Igh* V<sub>H</sub> inversion pro-B cells were separately purified via anti-B220 MicroBeads (Miltenyi, #130-049-501) from individual 4-6-week-old *Rag1* deficient mice (WT and *Igh* V<sub>H</sub> locus inversion) and were cultured in opti-MEM medium containing 10% (v/v) FBS plus 10 ng/mL IL-7 and 2 ng/mL SCF for 4-5 days as previously described<sup>38</sup>. Purified BM pro-B cells from different mice were separately cultured in different Petri dishes. WT and *Igh* V<sub>H</sub> inversion mice and short-term cultured BM-pro-B cells were double-checked and confirmed by PCR, respectively, prior to various assays as described below. PCR primers are listed in Supplementary Table 4.

### **Generation of doxycycline inducible Wapl-degrom v-*Ab1* pro-B lines**

The targeting plasmid (pTet-On\_OsTIR1\_V5) for introducing the doxycycline inducible OsTIR1-V5 expression cassette into endogenous *Rosa26* locus was constructed by modifying a

published pEN113 plasmid (Plasmid #86233, Addgene). The donor plasmid (pW290) used to target the endogenous mouse *Wapl* locus was constructed by modifying a published pMK290 plasmid (Plasmid #86230, Addgene). Briefly, the doxycycline inducible Tet-On-3G\_hPGK promoter sequence was PCR amplified from the plasmid pMK243 (Addgene, #72835) with 5' Sall and 3' MluI digestion sites. Then the pTet-On\_OsTIR1\_V5 targeting vector was generated by cutting out the enhancer and promoter sequence of OsTIR1\_V5 expression cassette from the pEN113 plasmid via Sall and MluI digestion, and replacing the amplified Tet-On-3G\_hPGK promoter sequence into pEN113 vector via the same restriction sites (Extended Data Fig.5b). To construct donor pW290 vector, 5' and 3' homology arms (247 bp each) flanking the BamHI sequence were synthesized via gBlock by IDT (Integrated DNA Technologies) and cloned into pGEM-T easy vector (Promega, #A1360) to generate the pGEM-Wapl plasmid. The mAID-mClover cassette containing a hygromycin selection marker was excised from pMK290 by BamHI and cloned at the BamHI site of the pGEM-Wapl plasmid, between the homology arms, to generate the pW290 vector (Extended Data Fig. 5d). To modify the mouse *Rosa26* and *Wapl* gene, sgRNAs targeting the endogenous *Rosa26* locus or 3' end sequence of the *Wapl* gene were cloned by annealing pairs of oligos into pX330 (Addgene, #42230) to construct the pX330\_*Rosa26* and pX330\_*Wapl*-mAID, respectively. Sequences of all sgRNAs and oligos used are listed in Supplementary Table 4.

The *Rag1*<sup>-/-</sup>; *Eμ-Bcl2*<sup>+</sup> *v-Abl* pro-B cells were co-transfected with pTet-On\_OsTIR1\_V5 and pX330\_*Rosa26* via Lonza 4D Nucleofector as shown before<sup>9,10</sup>, selected with 0.4 mg/ml of G418 for 5-6 days, and subcloned by dilution. Candidate clones with desired gene modifications were screened by PCR. The expression of OsTIR1 was confirmed by Western blot (Extended Data Fig. 5c). Similarly, the resultant clones were further co-transfected with donor pW290 with corresponding Cas9/sgRNA pX330\_*Wapl*-mAID via nucleofection, selected with 800 μg/ml hygromycin B (Millipore Sigma, #10843555001) for 3 days, and subcloned by dilution. Candidate clones with desired gene modifications were screened by PCR and confirmed by Southern blot in Extended Data Fig. 5d-e, and the expression of mAID-tagged Wapl protein was confirmed by Western blot and FACS as outlined in Extended Data Fig. 5f-j. The resulting lines were referred to as the *Wapl*-degron *v-Abl* pro-B lines and used in this study.

### **Treatment of *Wapl*-degron *v-Abl* pro-B lines with IAA&Dox**

To specifically deplete mAID-tagged Wapl protein in the G1 stage, *Wapl*-degron *v-Abl* pro-B lines were treated with 150 μM Indole-3-acetic acid (auxin analog, IAA, Sigma-Aldrich, #I3750-25G-A) and 2 mg/mL doxycycline (Dox, Sigma-Aldrich, #D3447) from day1 to day4 during 3 μM STI-571 (STI) induced four-day G1 arrest experimental period (day0-4) (Extended

Data Figure. 5f). All the chemicals were dissolved in DMSO to prepare the stock. DMSO solvent was applied as the mock to untreated cells. Both untreated and IAA&Dox-treated cells were then collected and examined by FACS for protein depletion confirmation (Extended Data Figure. 5h) prior to various assays as described below.

### **Western Blot**

The western blot experiment was performed according to published protocol with modification<sup>39</sup>. Before harvesting, the primary line with Tet-on OstIR1-V5 tag were treated with/without doxycycline for 24 hr and the Wapl-degrom *v-Ab1* cells were treated with/without IAA&Dox as described in Extended Data Fig. 5f. 10 million cells were harvested and lysed in EBC buffer (50 mM Tris-HCl pH 7.5, 120 mM NaCl, 0.5% NP40) with protease inhibitors (Complete Mini, Roche) for 30 min on ice, followed by pulse sonication for 5 s with low energy input. The lysates were then resolved by SDS-PAGE and immunoblotted with indicated antibodies (Wapl and V5-tag antibodies were purchased from Thermo Fisher Scientific #PA5-38024 and #R960-25, and  $\beta$ -actin antibody was purchased from Cell Signaling Technology #3700).

### **Cell viability assay**

For time-course cell viability assay, Wapl-degrom *v-Ab1* cells were treated with IAA&Dox from day1 to day4 during the STI-571 induced four-day G1 arrest experimental period (day0-4). The viability of Wapl-degrom *v-Ab1* pro-B cells with or without Wapl depletion at different time points (day 0 to day 4) was determined by the percentage of viable lymphocyte population gated by FACS side (SSC) and forward (FSC) scatters out of the total cells. Average percentage  $\pm$  s.d. of viable cells for each timepoint and each condition was calculated 4 independent experiments.

### **GFP quantification and cell cycle stage analysis**

For GFP (Clover) signaling quantification of Wapl depletion, both the untreated and IAA&Dox-treated Wapl-degrom *v-Ab1* cells at different time points (day0, day2 and day4) were directly collected and examined by FACS. Primary *v-Ab1* line #5 was processed as a GFP negative control.

For cell cycle analysis, Wapl-degrom *v-Ab1* pro-B cells without or with Wapl depletion at different time points (day0, day1, day2 and day4) were pelleted at 500 g for 5 min at 4 °C, washed with ice-cold PBS, pelleted again, then fixed in 66% ethanol on ice: gently resuspend the cell pellet in 400  $\mu$ L ice-cold 1 $\times$  PBS. Slowly add 800  $\mu$ L ice-cold 100% ethanol and mix well. Store at 4 °C for at least 2 hours. Gently resuspended the fixed cell pellet in 300  $\mu$ L 50  $\mu$ g/mL propidium iodide (PI) and 0.2 mg/mL RNase in PBS. After 30 min incubation in the dark

at 37 °C, cells were transferred on ice, passed through an appropriate filter to remove cell aggregates and then used directly for flow cytometry.

### **RAG complementation**

The *Rag1*-expressing vector pMSCV-RAG1-IRES-Bsr and *Rag2*-expressing vector pMSCV-FLAG-RAG2-GFP were kindly provided by Yu Zhang and Xuefei Zhang. RAG was reconstituted in RAG1-deficient *v-Abl* cells via retroviral infection of cells with the pMSCV-RAG1-IRES-Bsr and pMSCV-FLAG-RAG2-GFP vectors followed by 3-4 days of Blasticidin (ThermoFisher, #R21001) selection to enrich for cells with virus integration.

### **HTGTS-V(D)J-seq**

For *Igh* V(D)J recombination analyses, we purified BM pro-B cells from 4-6-week-old WT C57BL/6 mice and entire *Igh* V<sub>H</sub> locus inversion mice as described above and introduced RAG into *Rag1*<sup>-/-</sup> Wapl-degrom *v-Abl* pro-B cells as well as their derivatives via the approach described previously. HTGTS-V(D)J-seq libraries were prepared as previously described<sup>1,10,17,18</sup>. Briefly, 1 µg of gDNA from sorted mouse BM pro-B cells or 8 µg of gDNA from G1-arrested RAG-complemented Wapl-degrom *v-Abl* cells with or without Wapl depletion was sonicated and subjected to LAM-PCR using biotinylated J<sub>H</sub>1-4 bait primers. Single-stranded LAM-PCR products were purified using Dynabeads MyONE C1 streptavidin beads (Life Technologies, #65002) and ligated to bridge adaptors. Adaptor-ligated products were amplified by nested PCR with indexed J<sub>H</sub>1-4 primers and the primer annealed to the adaptor. The PCR products were further tagged with Illumina sequencing adaptor sequences, size-selected via gel extraction and loaded onto an Illumina NextSeq550 using paired-end 150-bp sequencing kit. Primer sequences are listed in Supplementary Table 4.

### **HTGTS-V(D)J-seq data processing and analyses**

HTGTS-V(D)J-seq libraries were processed via the pipeline described previously<sup>18</sup>. The data were aligned to mm9 genome and analyzed with all duplicate junctions included in the analyses as previously described. In the *Igh* repertoire libraries generated from C57BL/6 WT BM pro-B cells with J<sub>H</sub>1-4 coding end baits<sup>17,40</sup>, we detected in V<sub>H</sub>DJ<sub>H</sub> exons 109 functional V<sub>HS</sub>, as well as 23 pseudo V<sub>HS</sub> (Supplementary Table 1) (average rearrangement >1). For comparisons, utilization data of V<sub>H</sub> and D segments in WT or *Igh* V<sub>H</sub> inverted BM pro-B cells was normalized to 118,475 total recovered junctions (Fig. 1b; Supplementary Table 1). Utilization data of V<sub>H</sub> and D segments in G1-arrested Wapl-degrom *v-Abl* cells was normalized to 100,000 total recovered junctions (Fig. 3e; Extended Data Fig. 8b, c; Supplementary Table 2). Utilization data of V<sub>H</sub> and D segments in G1-arrested single *Igh* Wapl-degrom *v-Abl* cells with or without IGCR1 was normalized to 100,000 total recovered junctions (Extended Data Fig. 8b, e, f;

Supplementary Table 3). For the RAG off-targets in WT or *Igh* V<sub>H</sub> inverted BM pro-B cells, each library was isolated from individual independent J<sub>H</sub>1-4 primer libraries and was normalized to 402,241 isolated junctions (Fig. 2b; Extended Data Fig. 4). For comparison, each of the biological library replicates in BM pro-B cells was normalized to 2979 off-target junctions in the indicated region (chr12:114,666,712-121,257,530) for statistical analysis (Fig. 2c). For the RAG off-targets in Wapl-degrogen *v-Abl* cells and derived *Igh* V<sub>H</sub> inverted lines, each library was pooled from 3 independent J<sub>H</sub>1-4 primer libraries and was normalized to 432,676 isolated junctions without repetitive sequences (Extended Data Fig. 9b). For comparison, each of the biological library replicates in *v-Abl* lines was normalized to 500 off-target junctions in the indicated region (DQ52-entire *Igh* V<sub>H</sub>) for statistical analysis (Extended Data Fig. 9c). 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 (Fig. 2; Extended Data Fig. 9). Data reproducibility was controlled by performing at least three independent experiments and showed as mean ± s.d. for visualization.

### ChIP-Seq library preparation and analyses

ChIP-seq libraries were prepared with short-term cultured *Rag1*<sup>-/-</sup> BM pro-B cells or G1-arrested *Rag1*<sup>-/-</sup> *v-Abl* pro-B cells (with or without Wapl depletion) and performed based on previously description with modification<sup>2</sup>. In brief, 20 million cells were crosslinked in 37 °C prewarmed culture medium with 1% formaldehyde for 10 min at 37 °C and quenched with glycine at a final concentration of 125 mM. Cells were then treated with cell lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP40) for 10 min on ice, followed by treatment with nuclear lysis buffer (50 mM Tris-HCl pH 8.1, 10 mM EDTA, 1% SDS) for 10 min at room temperature. Chromatin was subjected to sonication with Diagenode Bioruptor at 4 °C to achieve an average size of 200-300 bp (30s on, 30 s off, 20 cycles with high energy input). Chromatin was then precleared with Dynabeads Protein A (Thermo Fisher Scientific, 10002D) at 4 °C for 2 h. 50 mL lysates was kept as input and the rest were incubated with 5 µg RAD21 antibody (Abcam, #ab992) or 10 µg CTCF antibody (Millipore, #07-729) or 5 µg Wapl antibody (Thermo Fisher Scientific, #PA5-38024) overnight at 4 °C. Immunoprecipitation (IP) samples were then captured by Dynabeads Protein A at 4 °C for at least 2 h 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 NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, #E7645) and sequenced by paired-end 75-bp sequencing on Next-Seq550. For comparison of all the ChIP-seq results in this study, we used

bowtie2 to align ChIP-seq reads to mm9 genome and run MACS2 callpeak with parameters '-t IP.bam -c input.bam -nomodel -keep-dup all -extsize 51 -nolambda -B -SPMR -g mm -broad' to generate bigwig graph normalized to 10 million reads. For comparison of ChIP-seq results between WT and *Igh* inversion BM pro-B cells, Rad21 ChIP-seq profiles across the 4 V<sub>H</sub> domains were potted in the normal orientation (Extended Data Fig. 3a-c). Indicated replicates were combined as mean ± s.d. for visualization.

### GRO-seq and data analyses

GRO-seq libraries were prepared as described previously<sup>1,41</sup> on short-term cultured *Rag1*<sup>-/-</sup> BM pro-B cells and G1-arrested *Rag1*<sup>-/-</sup> Wapl-degrogen *v-Abl* cells with or without Wapl depletion. Briefly, 10 million cells were collected and permeabilized with the DEPC treated buffer (10 mM Tris-HCl pH 7.4, 300 mM sucrose, 10 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM EGTA, 0.05% Tween-20, 0.1% NP40 substitute, 0.5 mM DTT, protease inhibitors and RNase inhibitor). The permeabilized cells were resuspended in 100 µl of DEPC treated storage buffer (10 mM Tris-HCl pH 8.0, 25% (v/v) glycerol, 5 mM MgCl<sub>2</sub>, 0.1 mM EDTA and 5 mM DTT) followed by nuclear run-on with 100 µl 2X run-on mix (5 mM Tris-HCl pH 8.0, 2.5 mM MgCl<sub>2</sub>, 0.5 mM DTT, 150mM KCl, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 0.5 mM Br-UTP, RNase inhibitor, 1% Sarkosyl) at 37 °C for 5 min. Total RNA was extracted by Trizol and followed by hydrolyzation with NaOH at a final concentration of 0.2 N on ice for 18 min. After quenching with ice-cold Tris-HCl pH 6.8 at a final concentration of 0.55 M and exchanging buffer via Bio-Rad P30 columns, the total RNA was incubated with Br-dU antibody-conjugated beads (Santa Cruz, #sc-32323-ac) for 1 h. The enriched Br-dU labeled RNAs were incubated with RppH (NEB, #M0356S) and with T4 PNK (NEB, #M0201S) for hydroxyl repair, followed by ligating the 5' and 3' RNA adaptors. RT-PCR was performed after adaptor ligation to obtain cDNAs. Half of the cDNAs was subjected to library preparation by two rounds of PCR with barcoded primers. The second round of PCR products were purified by AMPure beads (Beckman Coulter, #A63880). GRO-seq libraries were sequenced via paired-end 75 bp sequencing on Illumina NextSeq550. For comparison of all the GRO-seq results in this study, we used bowtie2 to align GRO-seq reads to mm9 genome and run MACS2 callpeak with parameters '-nomodel -keep-dup all -extsize 51 -nolambda -B -SPMR -g mm -broad' to generate bigwig graph. To visualize the genome-wide RNA expression level, we run htseq-count to count read number on each gene and made scatter plot after down-sampling each sample to 10 million reads mappable to any genes. For comparison of GRO-seq results between WT and *Igh* inversion BM pro-B cells, GRO-seq profiles across the 4 V<sub>H</sub> domains were potted in the normal orientation. Both the sense and antisense transcription are relative to the entire *Igh* V<sub>H</sub> locus upstream V<sub>H</sub>81X with or without



inversion and indicated respectively (Extended Data Fig. 3d). Spearman's correlation coefficient ( $\rho$ ) and  $p$  values determined by two-sided Spearman's correlation test are presented between indicated samples.  $p$  values are calculated and shown in the figures as the follows: non-significant (NS):  $p \geq 0.05$ , \*:  $0.01 \leq p < 0.05$ , \*\*:  $0.001 \leq p < 0.01$ , and \*\*\*:  $p < 0.001$ .

### 3C-HTGTS and data analyses

3C-HTGTS on short-term cultured *Rag1*<sup>-/-</sup> BM pro-B cells, G1-arrested Wapl-degrom *v-Abl* cells without or with IAA&Dox treatment, was performed as previously described<sup>10</sup>. Briefly, 10 million cells were crosslinked with 2% formaldehyde (Sigma-Aldrich, #F8775) for 10 min at room temperature and quenched with glycine at a final concentration of 125 mM. Cells were lysed on ice for 10 min followed by centrifugation to get nuclei. Nuclei were resuspended in NEB Cutsmart buffer for NlaIII (NEB, #R0125) digestion at 37 °C overnight, followed by T4 DNA ligase (Promega, #M1801) mediated ligation under dilute conditions at 16 °C overnight. Ligated products were treated with Proteinase K (Roche, #03115852001) and RNase A (Invitrogen, #8003089) followed by DNA purification to get the 3C templates. 3C-HTGTS library preparation follows the standard LAM-HTGTS library preparation procedures as previously described<sup>10</sup>. 3C-HTGTS libraries were sequenced via Illumina NextSeq550 using paired-end 150-bp sequencing kit. For comparison, sequencing reads of all the 3C-HTGTS libraries were aligned to mm9 genome and processed as previously described<sup>10</sup>. Data were plotted for comparison after normalizing junction from each experimental 3C-HTGTS library by random selection to the total number of genome-wide or entire *Igh* locus ( $\pm 10$  kb, chr12:114,444,081-117,258,165) junctions recovered from the smallest library in the set of libraries being compared. Chromosomal interaction patterns were very comparable before and after normalization. For short-term culture *Rag1*<sup>-/-</sup> BM pro-B cells, 3C-HTGTS libraries using iE $\mu$  and V<sub>H</sub>8-7P (Fig. 1c, Extended Data Fig. 2) as baits were normalized to 885,651 and 742,577 total junctions of genome-wide, respectively. For comparison of iE $\mu$ -baited interactions between WT and *Igh* inversion BM pro-B cells in the inverted orientation, 3C-HTGTS results were inverted in the *Igh* inversion BM pro-B cells (Fig. 1c, bottom). For G1-arrested Wapl-degrom *v-Abl* cells with or without Wapl depletion, 3C-HTGTS libraries using V<sub>H</sub>1-47 (Extended Data Fig. 7a) as baits were normalized to 231,542 total junctions of entire *Igh* locus ( $\pm 10$  kb). For comparison of 3C-HTGTS libraries between short-term culture *Rag1*<sup>-/-</sup> BM pro-B cells and G1-arrested Wapl-degrom *v-Abl* cells with or without Wapl depletion, we removed the reads in the regions closely near to the baits (iE $\mu$  bait, chr12:114,664,385-114,668,233; V<sub>H</sub>8-7P bait, chr12:116,495,497-116,495,577), which did not contribute to the locus interaction. By this way, 3C-HTGTS libraries using iE $\mu$  (Fig. 3d; Extended Data Fig. 8d) and V<sub>H</sub>8-7P (Extended Data Fig. 7b) as baits were normalized to 53,795 and

195,259 total junctions of entire *Igh* locus ( $\pm 10$  kb), respectively. The sequences of primers used for generating 3C-HTGTS libraries are listed in Supplementary Table 4.

### Data availability

HTGTS-V(D)J-Seq, 3C-HTGTS, CHIP-seq, and GRO-seq sequencing data reported in this study have been deposited in the GEO database under the accession number GSE151910. Specifically, HTGTS-V(D)J-Seq data related to Fig. 1b, 2, 3e-f and Extended Data Fig. 4, 8b-c, 8e-f, 9 are deposited in the GEO database under the accession number GSE151906. 3C-HTGTS data related to Fig. 1c, 3d, and Extended Data Fig. 2, 7, 8d are deposited in the GEO database under the accession number GSE151909. CHIP-seq data related to Extended Data Fig. 3a-c, 6a-b, 6e are deposited in the GEO database under the accession number GSE151908. GRO-seq data related to Fig. 3a-b and Extended Data Fig. 3d, 6c-d are deposited in the GEO database under the accession number GSE151907.

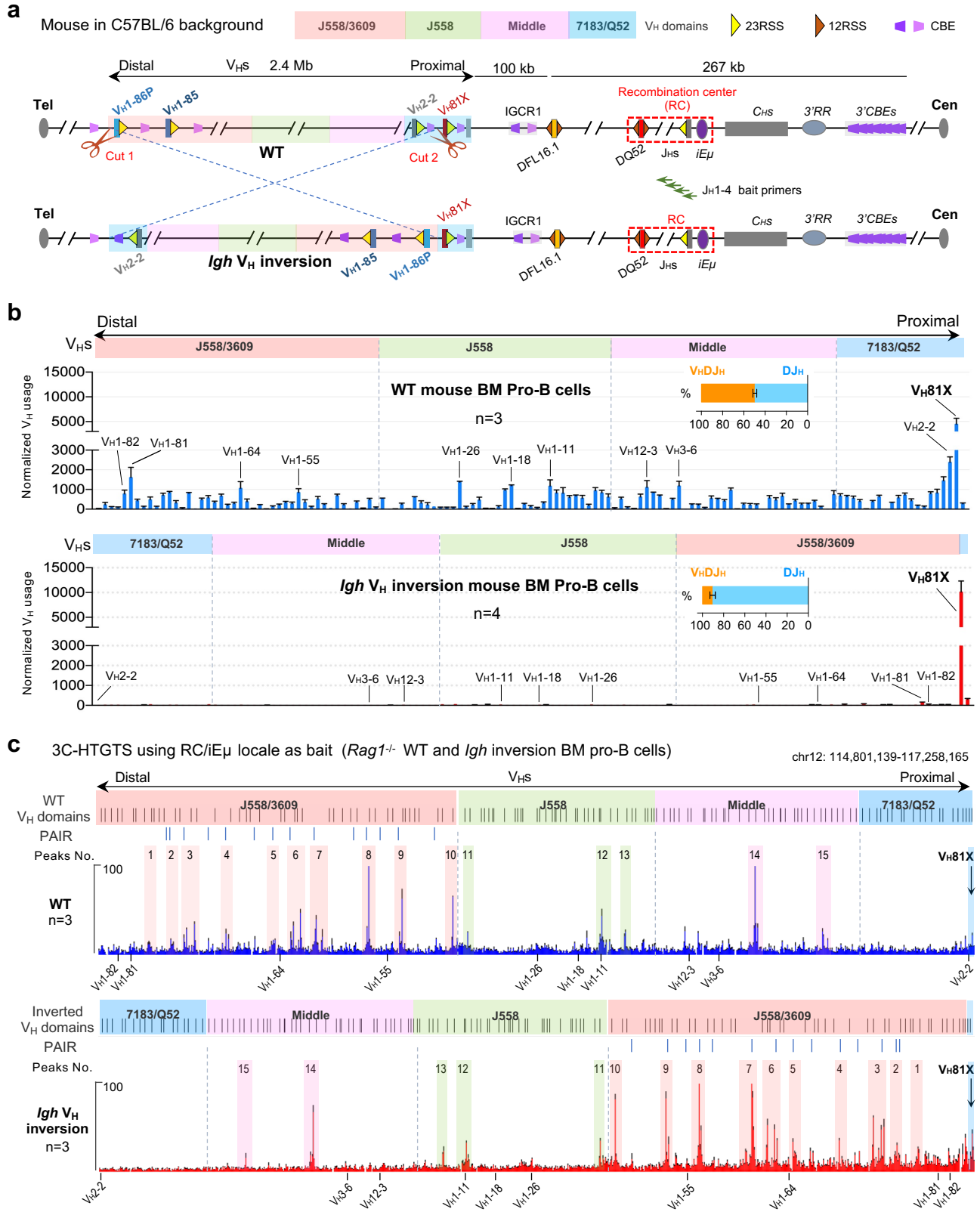
### Code availability

HTGTS-V(D)J-seq, 3C-HTGTS, CHIP-seq, and GRO-seq data were processed through the published pipelines as previously described<sup>18</sup>. Specifically, these pipelines are available at [http://robinmeyers.github.io/transloc\\_pipeline/](http://robinmeyers.github.io/transloc_pipeline/) (HTGTS pipeline), [http:// bowtie-bio.sourceforge.net/bowtie2/index.shtml](http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) (Bowtie2 v.2.2.8), <https://sourceforge.net/projects/samtools/files/samtools/1.8/> (SAMtools v.1.8)

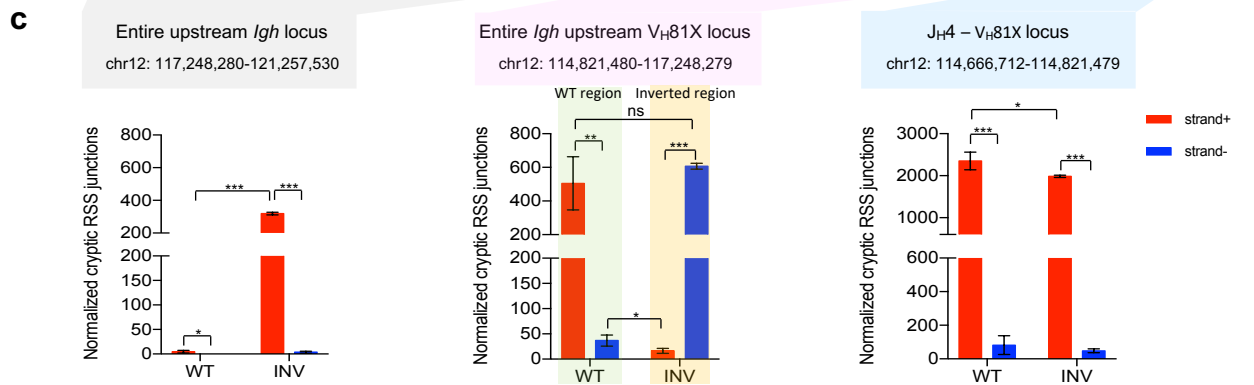
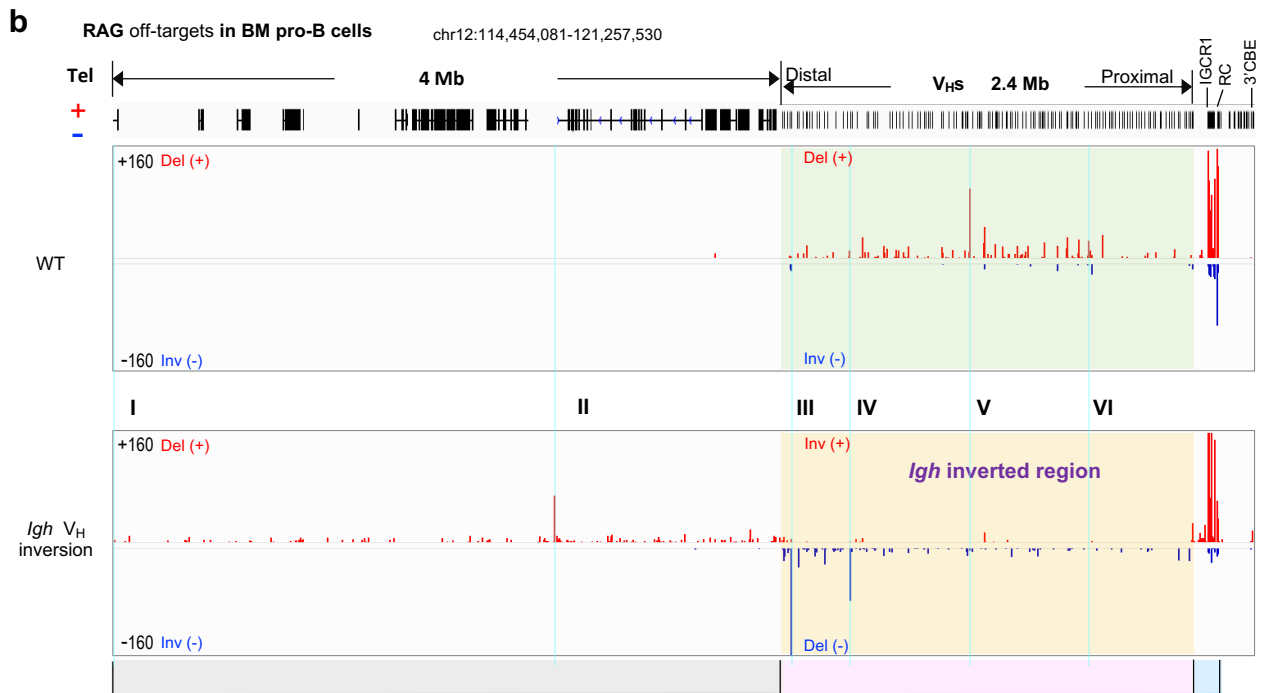
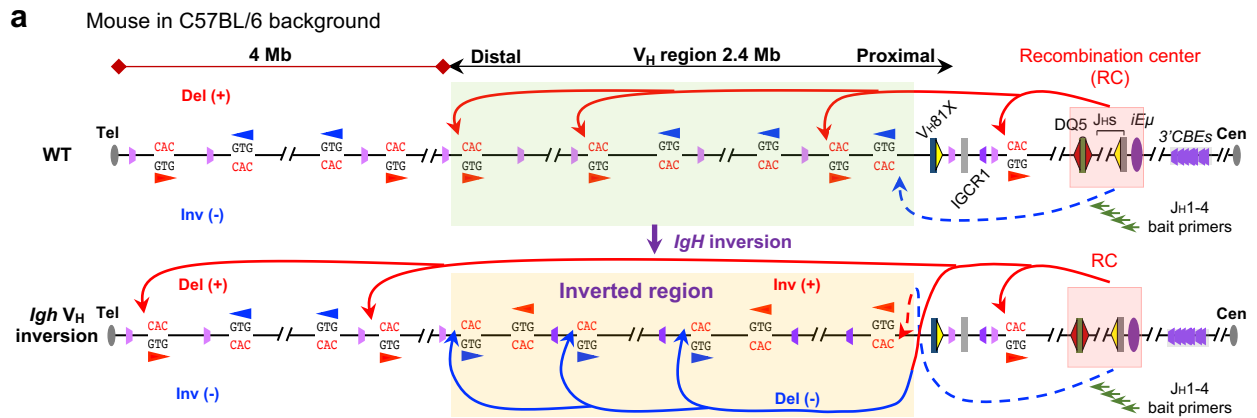
- 36 Dai H-Q, Liang. Z., Chang AN, Chapdelaine-Williams AM, Alvarado B, Pollen A, Alt FW and Schwer B. Direct analysis of brain phenotypes via neural blastocyst complementation. *Nature Protocols*. *In press* (2020).
- 37 Chen, J., Lansford, R., Stewart, V., Young, F. & Alt, F. W. RAG-2-deficient blastocyst complementation: an assay of gene function in lymphocyte development. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 4528-4532, doi:10.1073/pnas.90.10.4528 (1993).
- 38 Sayegh, C. E., Jhunjunwala, S., Riblet, R. & Murre, C. Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes & development* **19**, 322-327, doi:10.1101/gad.1254305 (2005).
- 39 Tseng, A., Inuzuka, H., Gao, D., Singh, A. & Wei, W. Experimental approaches to investigate the proteasomal degradation pathways involved in regulation of apoptosis. *Methods in enzymology* **446**, 205-223, doi:10.1016/S0076-6879(08)01612-1 (2008).
- 40 Chen H, Zhang. Y., Ye A, Du, Z, Xu M, Lee C-S, Hwang JK, Kyritsis N, Ba Z, Neuberg D, Littman DR & Alt FW. BCR Selection and Affinity Maturation in Peyer's Patches Germinal Centers. *Nature*, 582, pages421–425 (2020)
- 41 Zhang X, Zhang Y, Ba Z, Kyritsis N, Casellas R, Alt FW. Fundamental roles of chromatin loop extrusion in antibody class switching. *Nature* 575, 385-389, doi:10.1038/s41586-019-1723-0 (2019).



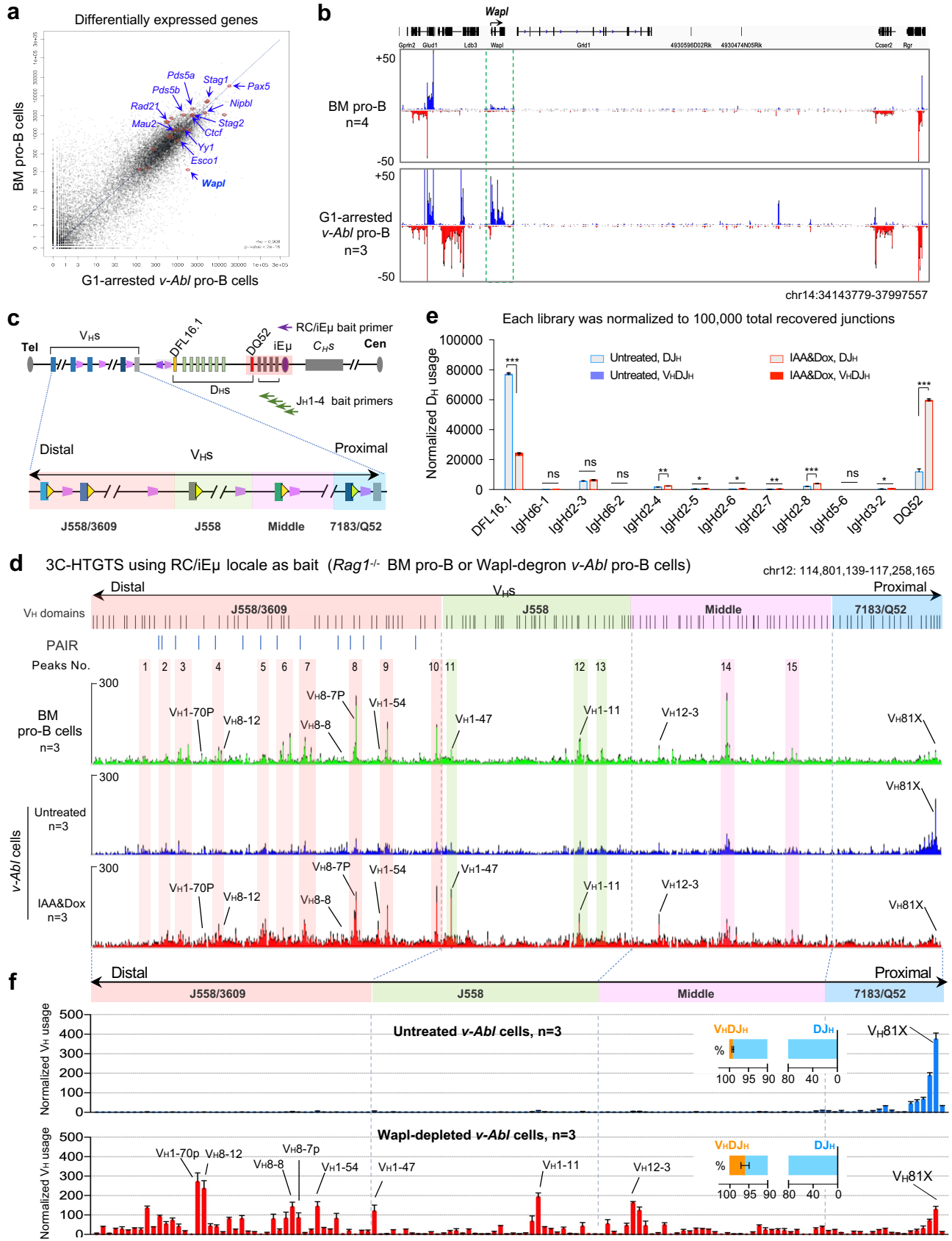
# Figure. 1



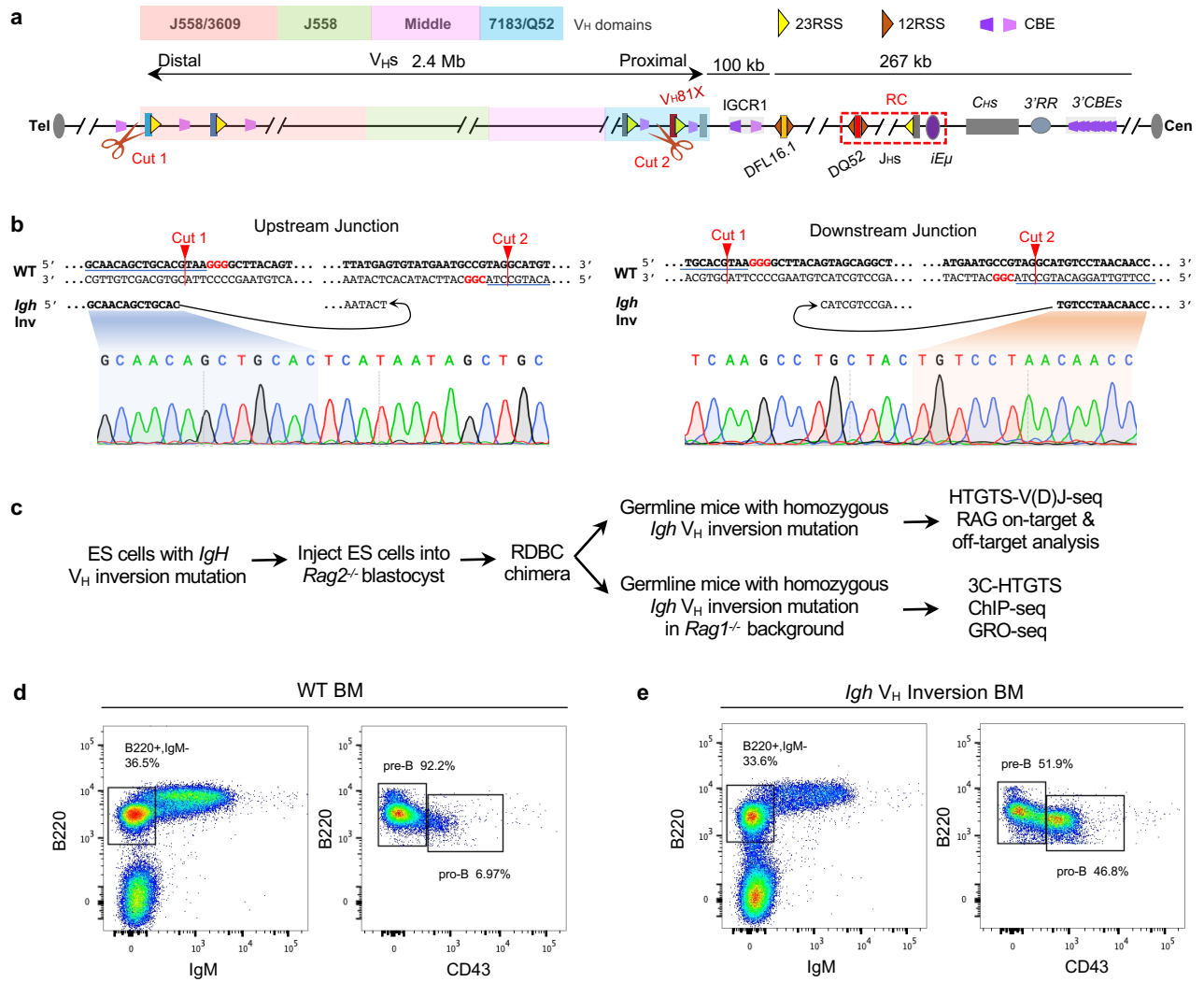
# Figure. 2



# Figure. 3

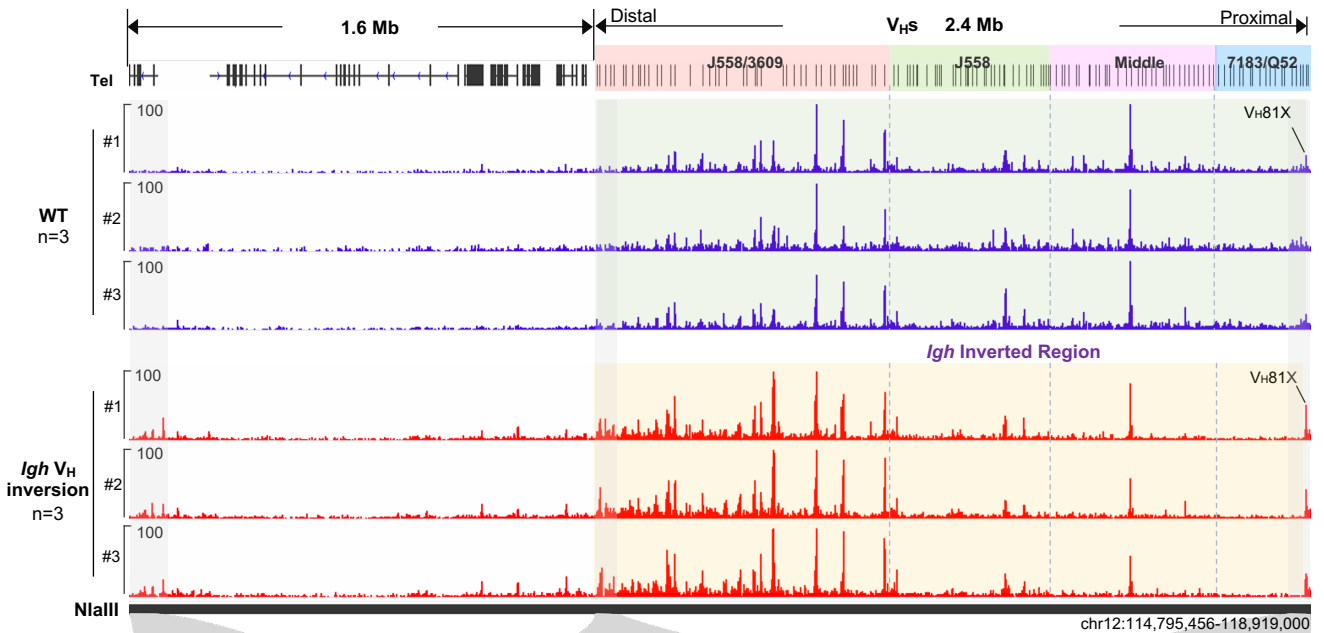


# Extended Data Fig. 1

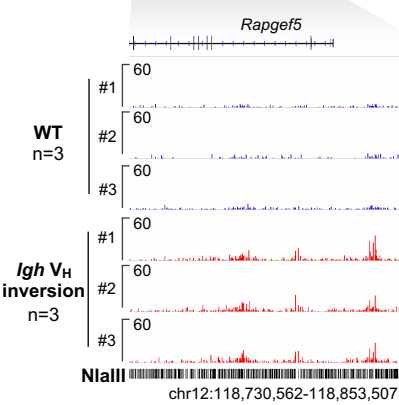


## Extended Data Fig. 2

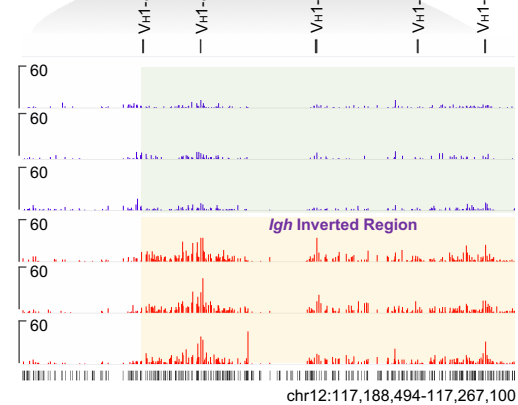
**a** 3C-HTGTS using RC/iEμ locale as bait (*Rag1*<sup>-/-</sup> WT and *Igh* inversion BM pro-B cells)



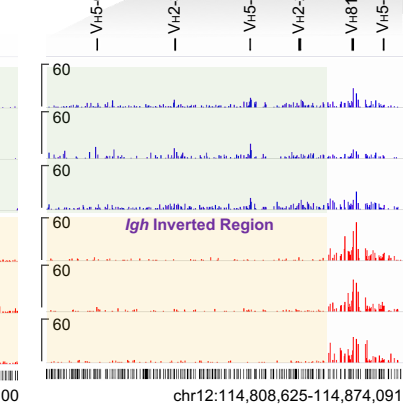
**b**



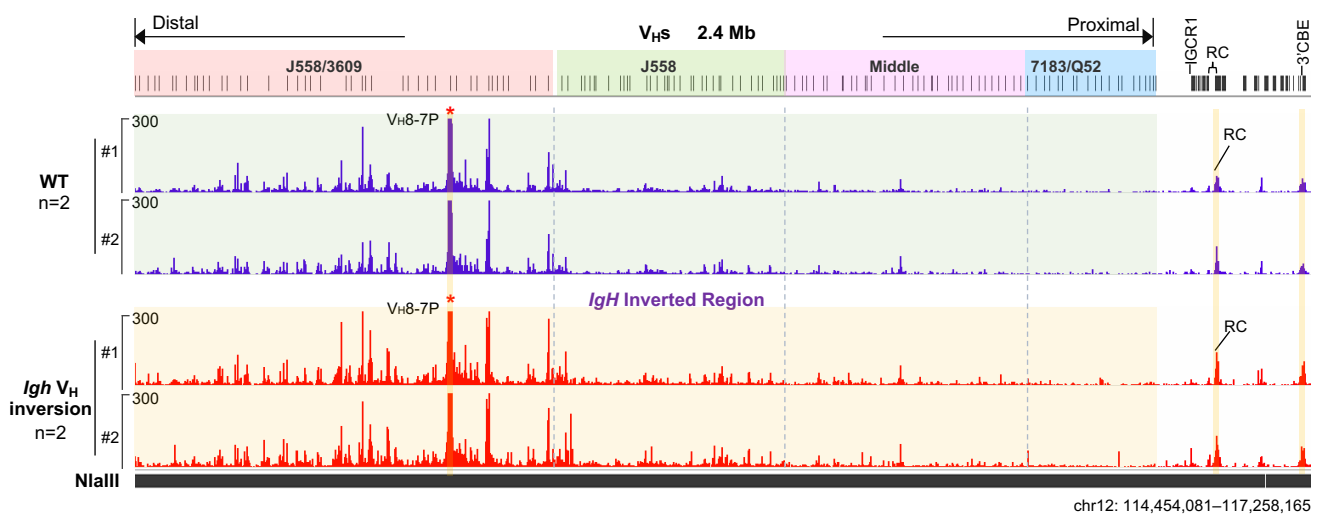
**c**



**d**

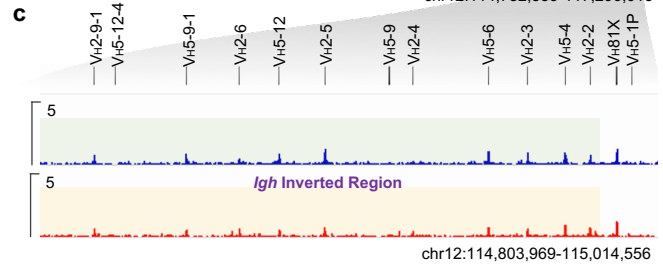
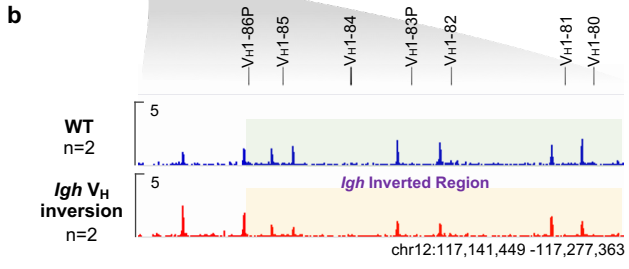
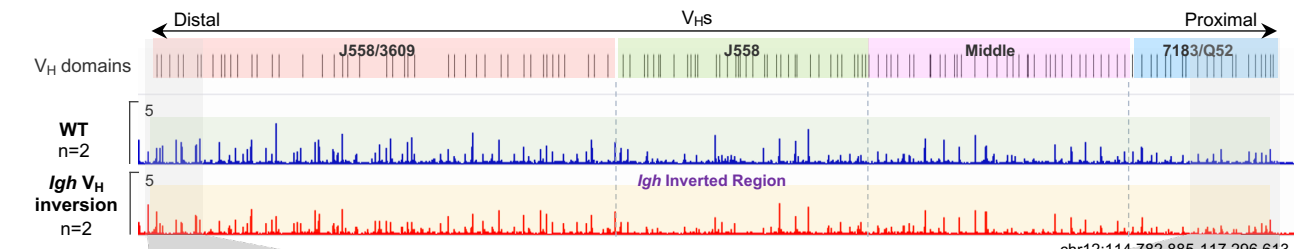


**e** 3C-HTGTS using V<sub>H</sub>8-7P\* locale as bait (*Rag1*<sup>-/-</sup> WT and *Igh* inversion BM pro-B cells)

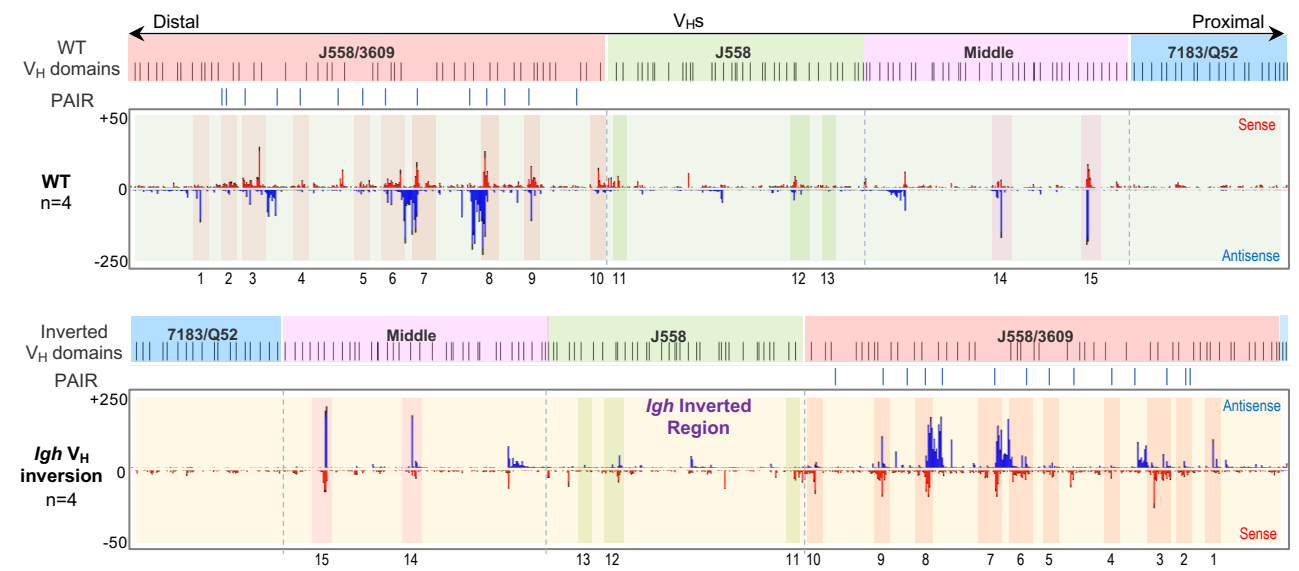


### Extended Data Fig. 3

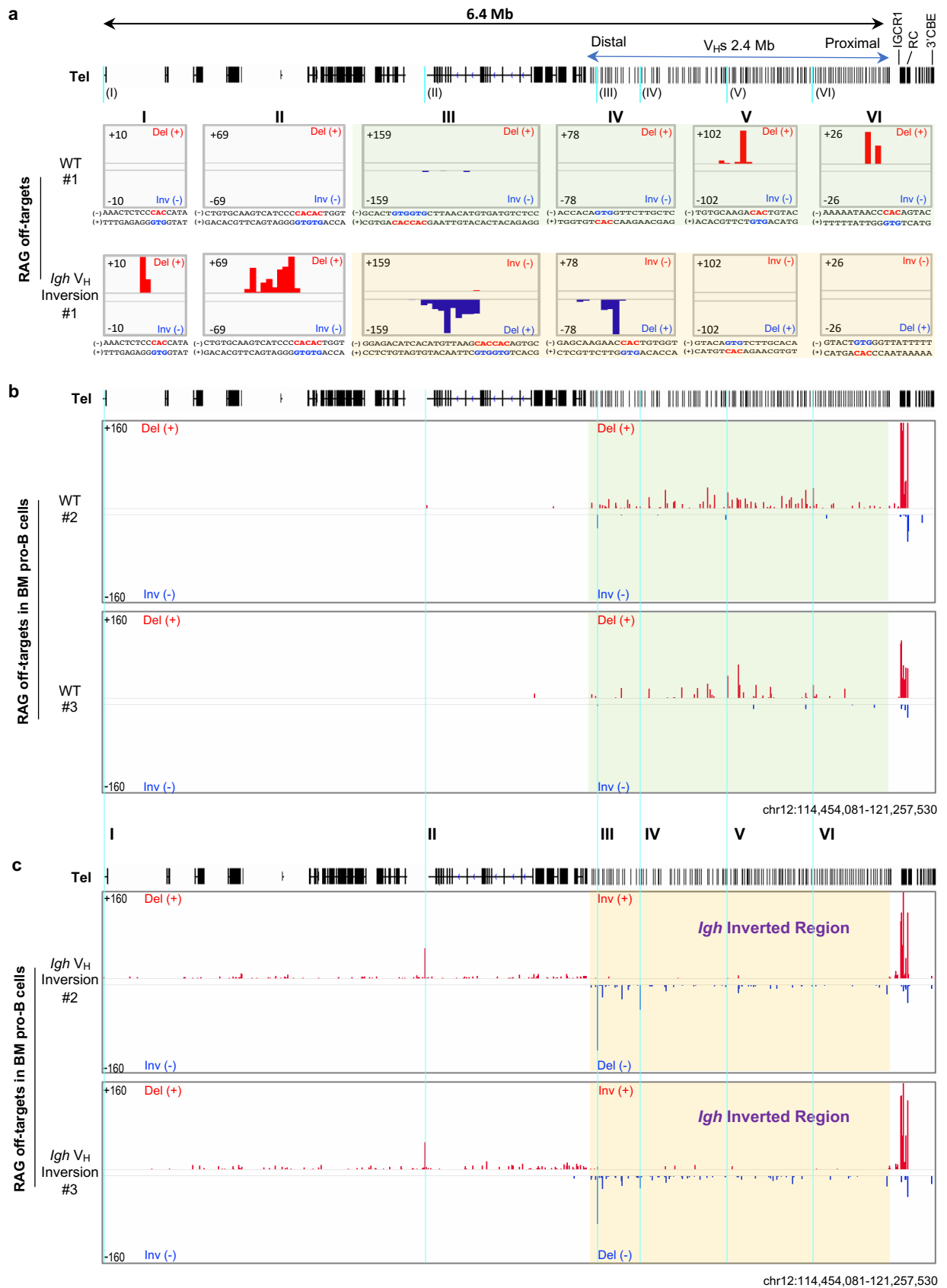
**a** Rad21 ChIP-seq profiles at *Igh* V<sub>H</sub> locus (*Rag1*<sup>-/-</sup> WT and *Igh* V<sub>H</sub> inversion BM pro-B cells)



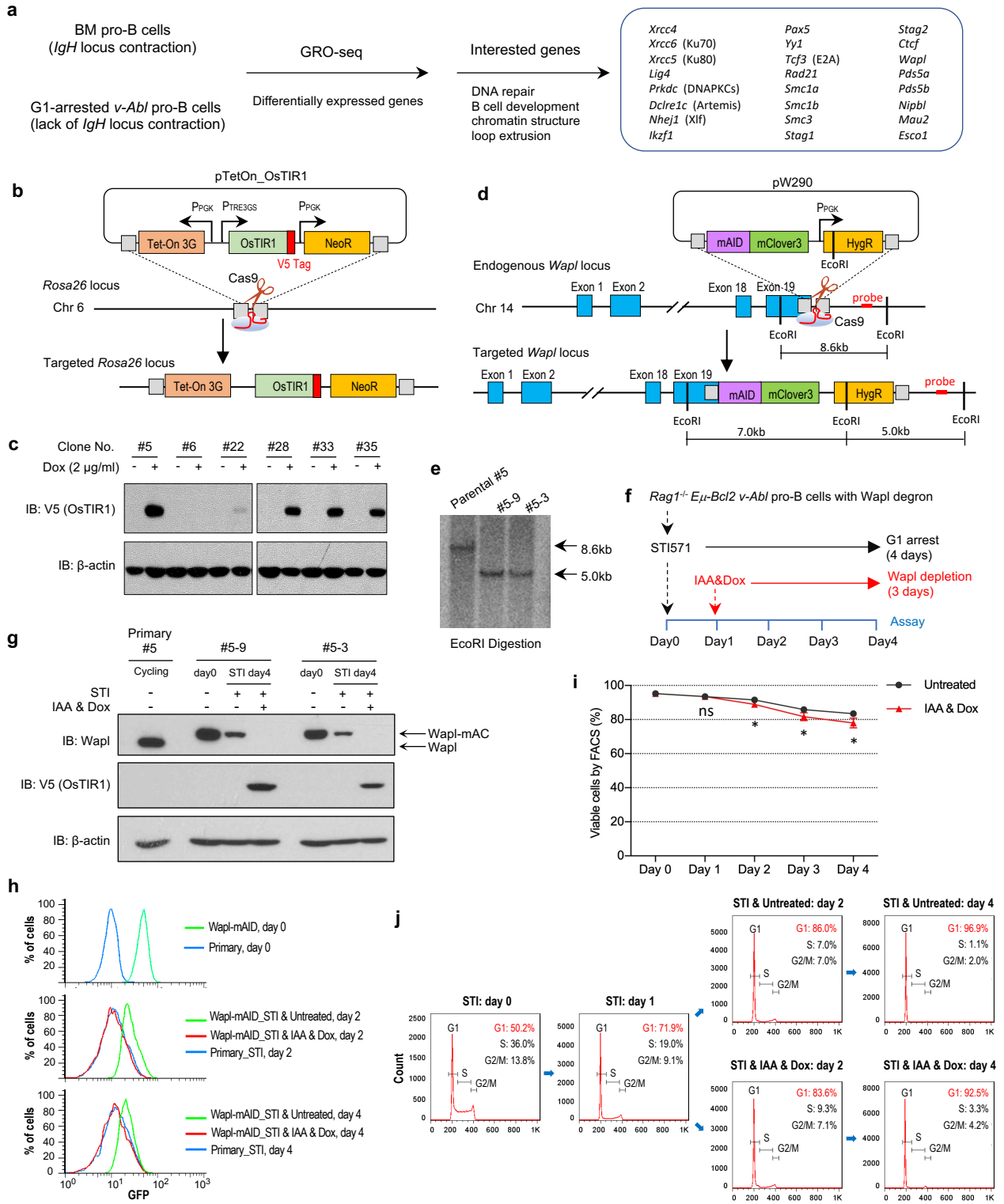
**d** GRO-seq profiles at *Igh* V<sub>H</sub> locus (*Rag1*<sup>-/-</sup> WT and *Igh* V<sub>H</sub> inversion BM pro-B cells)



# Extended Data Fig. 4



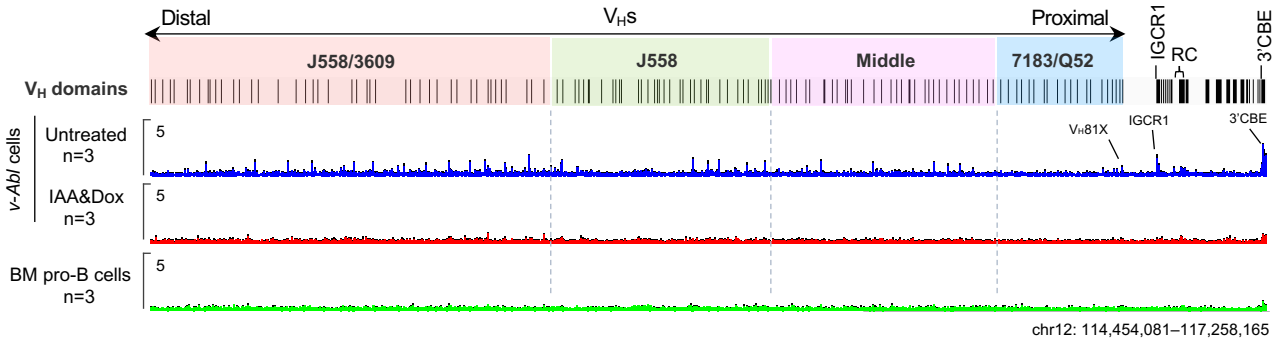
# Extended Data Fig. 5



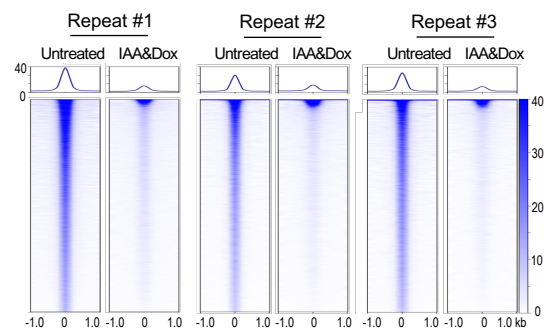


# Extended Data Fig.6

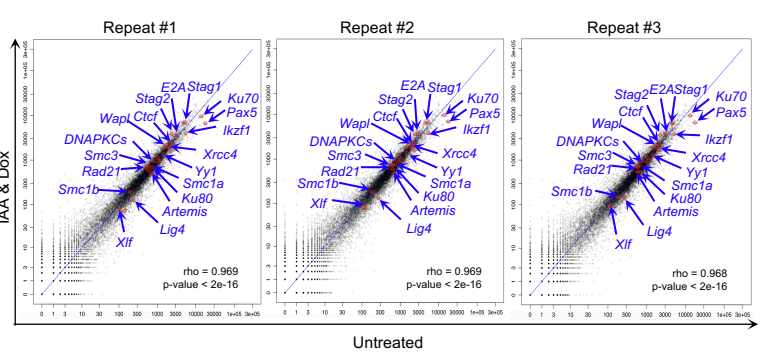
## a Wapl ChIP-seq profiles at *Igh* locus



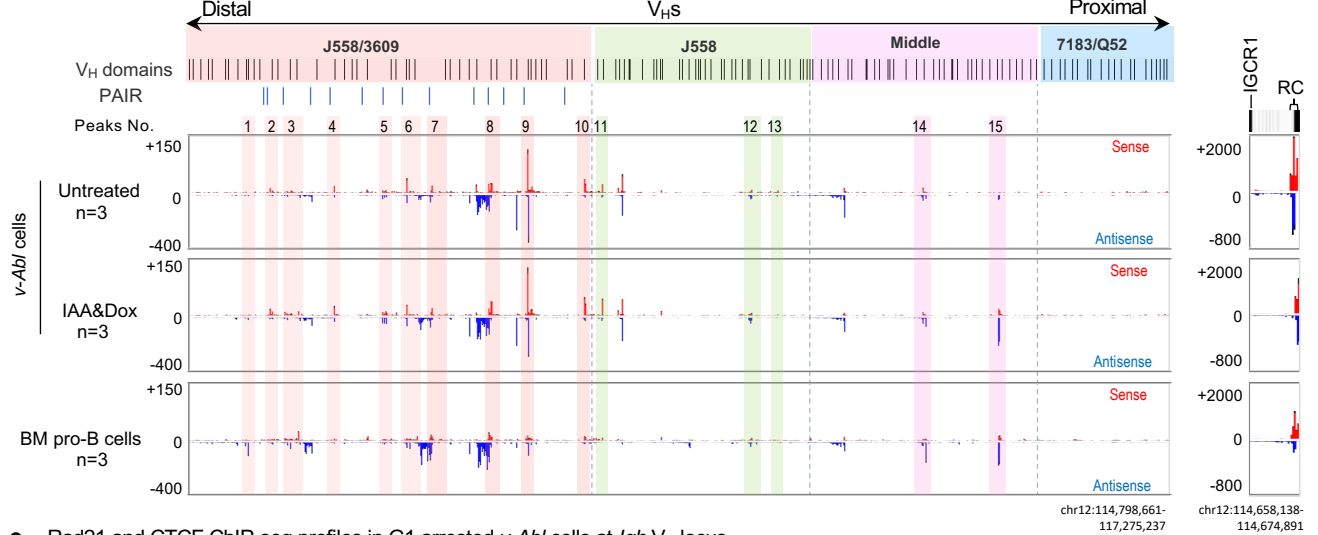
## b Genome-wide Wapl ChIP-seq peaks in G1 arrested *v-Ab/* cells



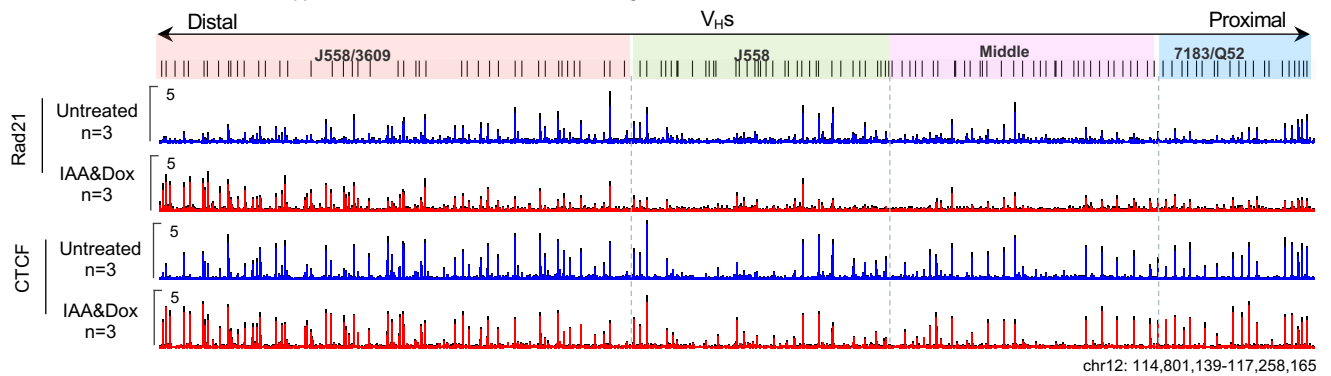
## d Genome-wide GRO-seq analyses in G1 arrested *v-Ab/* cells



## c GRO-seq profiles at *Igh* locus

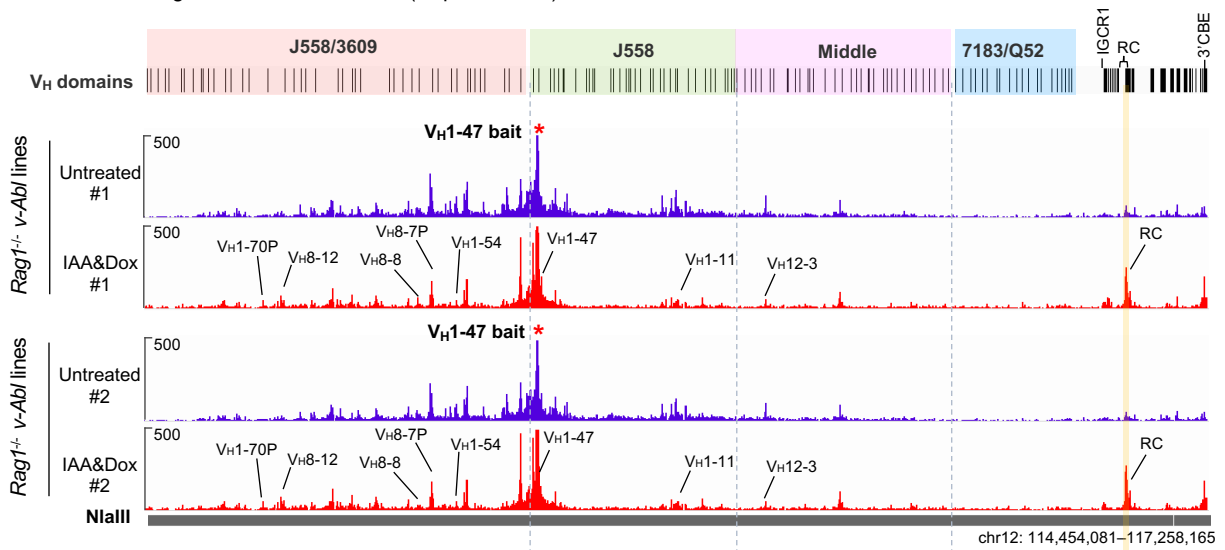


## e Rad21 and CTCF ChIP-seq profiles in G1 arrested *v-Ab/* cells at *Igh* $V_H$ locus

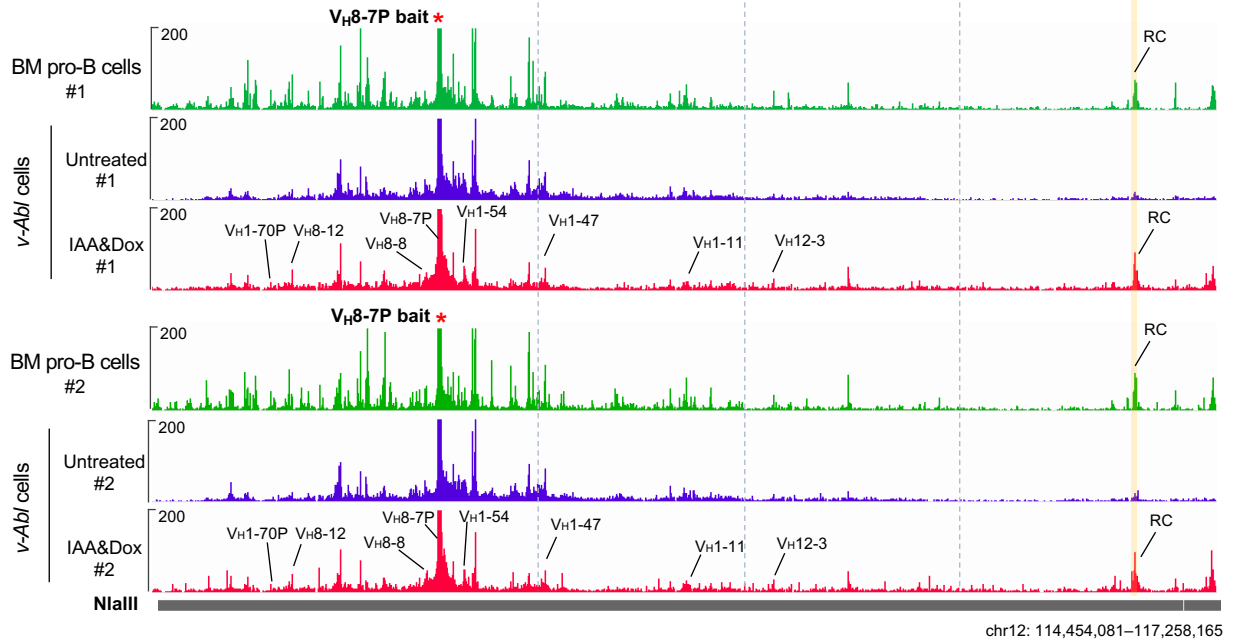


# Extended Data Fig. 7

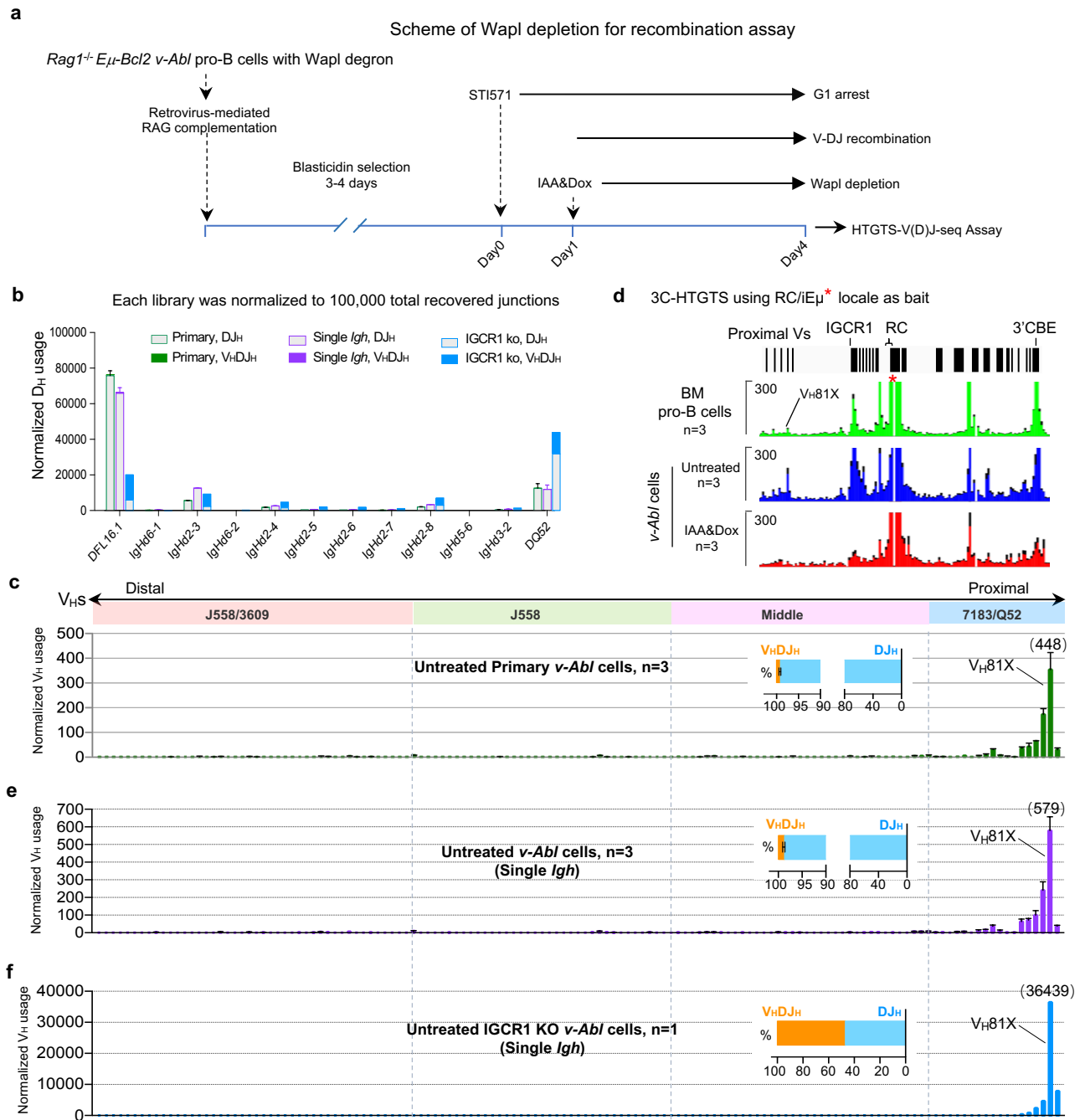
**a** 3C-HTGTS using V<sub>H</sub>1-47\* locale as bait (Repeat #1, #2)



**b** 3C-HTGTS using V<sub>H</sub>8-7P\* locale as bait (Repeat #1, #2)



# Extended Data Fig. 8



# Extended Data Fig. 9

