Molecular Cell, Volume *63*

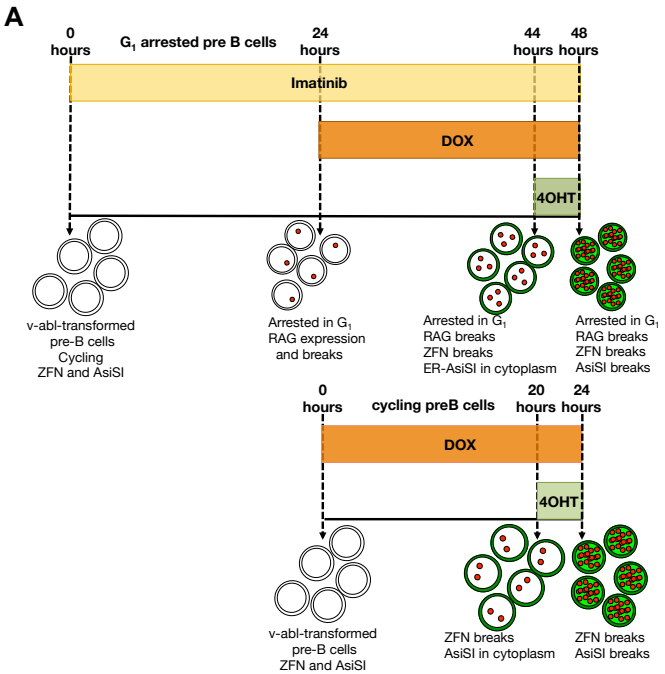
Supplemental Information

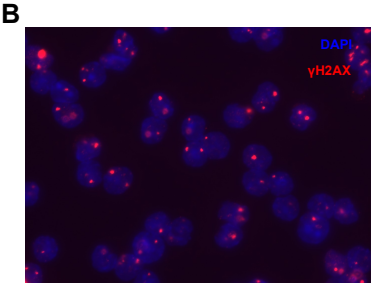
DNA Breaks and End Resection Measured

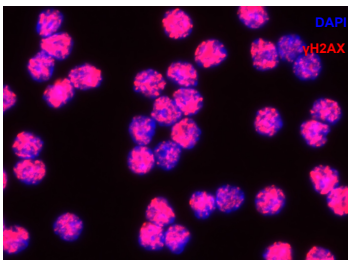
Genome-wide by End Sequencing

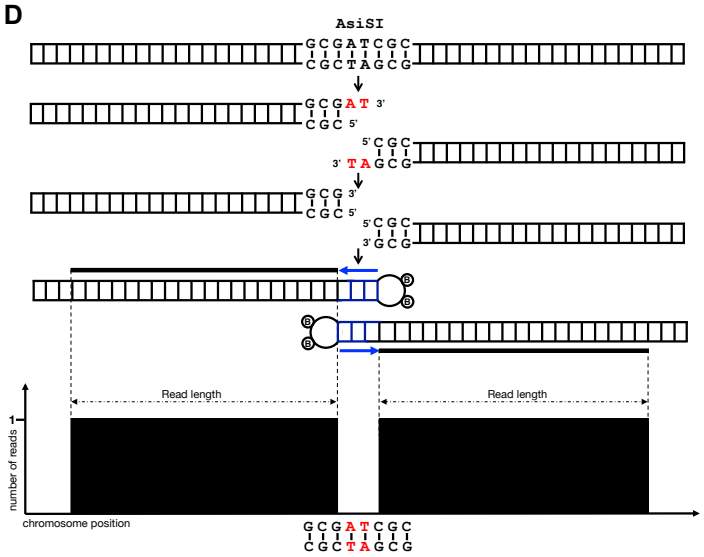
Andres Canela, Sriram Sridharan, Nicholas Sciascia, Anthony Tubbs, Paul Meltzer, Barry P. Sleckman, and André Nussenzweig

Figure S1


A
0 hours
G1 arrested pre B cells
24 hours
44 hours
48 hours
Imatinib
DOX
4OHT
v-abl-transformed pre-B cells Cycling ZFN and AsiSI
Arrested in G1 RAG expression and breaks
Arrested in G1 RAG breaks ZFN breaks ER-AsiSI in cytoplasm
Arrested in G1 RAG breaks ZFN breaks AsiSI breaks
0 hours
cycling preB cells
20 hours
24 hours
DOX
4OHT
v-abl-transformed pre-B cells ZFN and AsiSI
ZFN breaks AsiSI in cytoplasm
ZFN breaks AsiSI breaks


B
DAPI
γH2AX


C
DAPI
γH2AX


D
AsiSI
Read length
Read length
number of reads
1
chromosome position

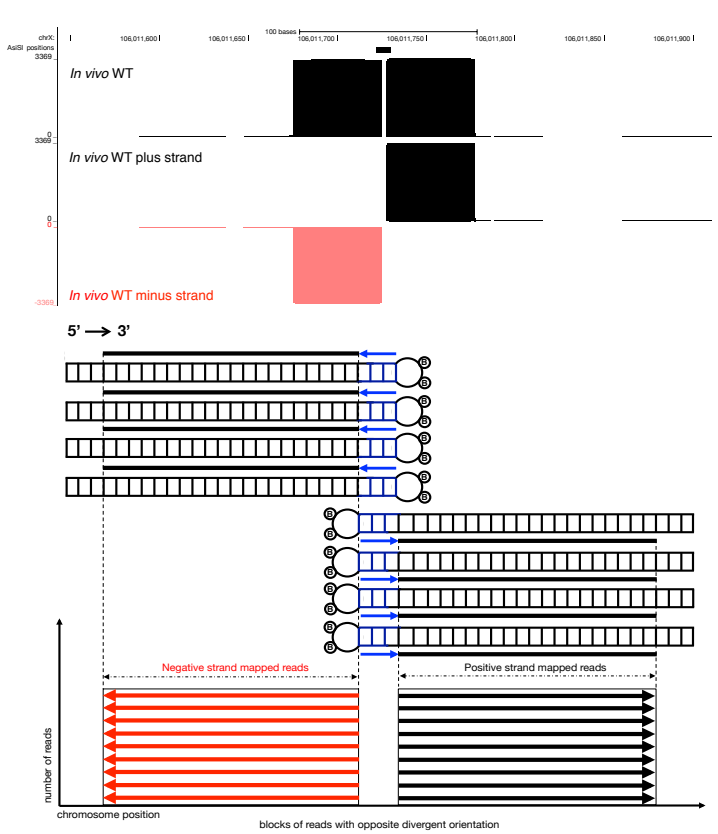
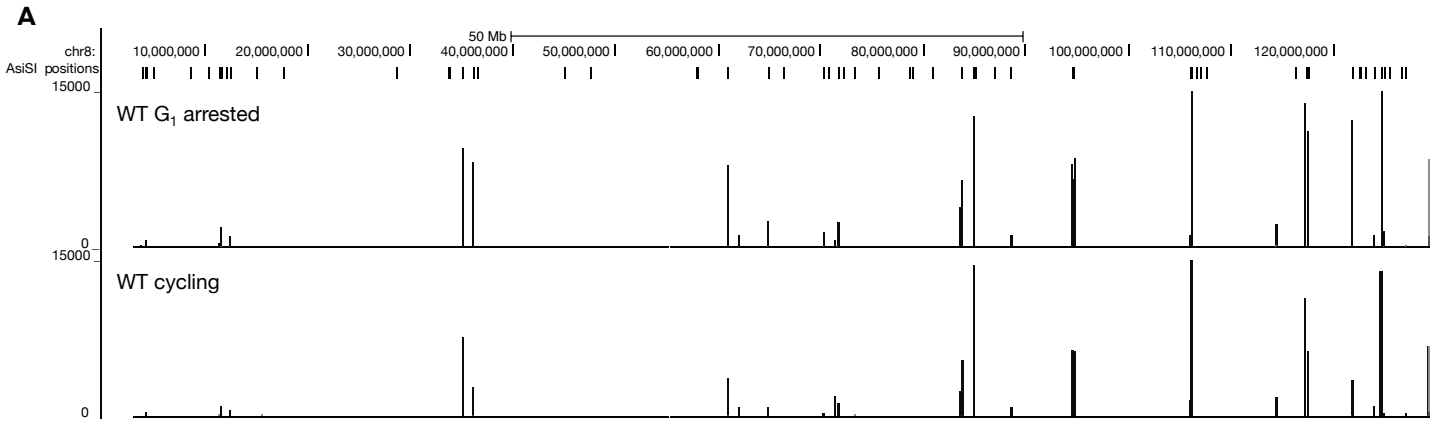
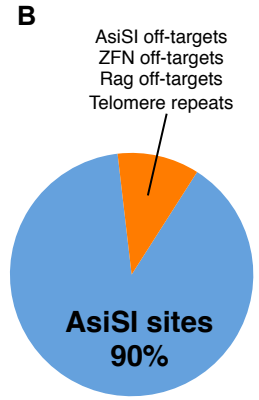
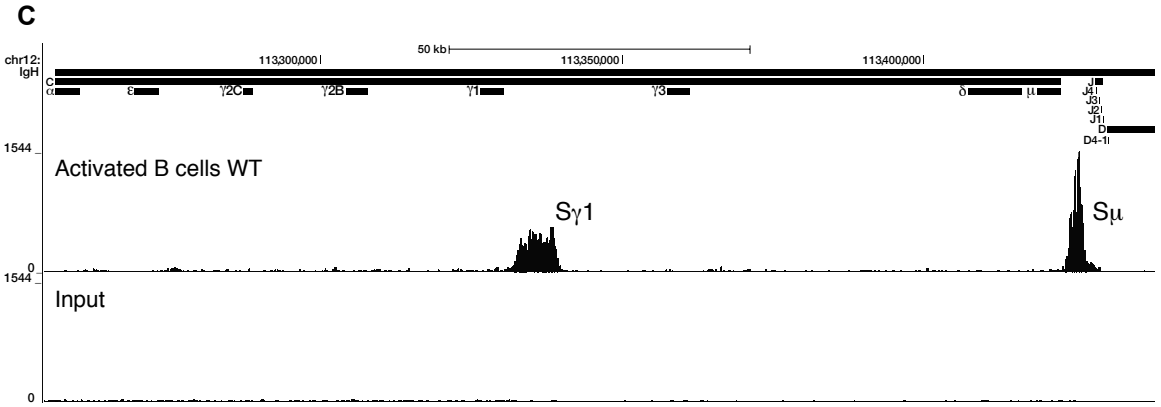
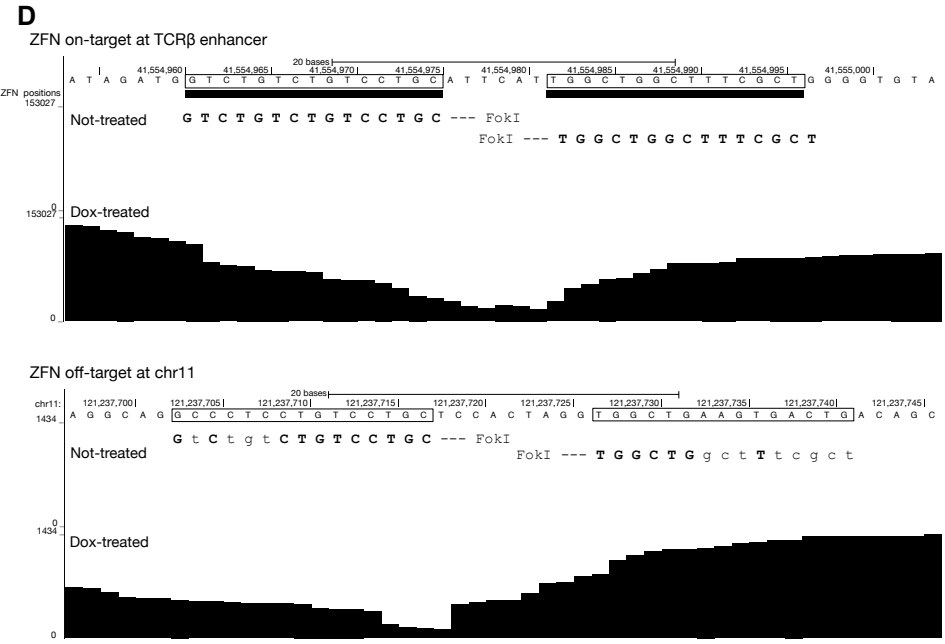

E
In vivo WT
In vivo WT plus strand
In vivo WT minus strand
5' → 3'
Negative strand mapped reads
Positive strand mapped reads
number of reads
chromosome position
blocks of reads with opposite divergent orientation

Figure S2


A
chr8:
AsiSI positions
50 Mb
10,000,000
20,000,000
30,000,000
40,000,000
50,000,000
60,000,000
70,000,000
80,000,000
90,000,000
100,000,000
110,000,000
120,000,000
15000
WT G1 arrested
0
15000
WT cycling
0


B
AsiSI off-targets
ZFN off-targets
Rag off-targets
Telomere repeats
AsiSI sites
90%


C
chr12:
IgH
50 kb
113,300,000
113,350,000
113,400,000
Cα
ε
γ2C
γ2B
γ1
γ3
δ
μ
Jμ
Jε1
Jε2
Jε3
Jε4
D
D4-1
1544
Activated B cells WT
Sγ1
Sμ
0
1544
Input
0


D
ZFN on-target at TCRβ enhancer
20 bases
41,554,960
41,554,965
41,554,970
41,554,975
41,554,980
41,554,985
41,554,990
41,554,995
41,555,000
ATAGATGGTCTGTCTGTCCTGCATTCATTGGCTGGCTTTCGCTGGGGTGTA
ZFN positions
153027
Not-treated
GTCTGTCTGTCCTGC --- FokI
FokI --- TGGCTGGCTTTCGCT
Dox-treated
0
153027
ZFN off-target at chr11
20 bases
chr11:
121,237,700
121,237,705
121,237,710
121,237,715
121,237,720
121,237,725
121,237,730
121,237,735
121,237,740
121,237,745
AGGCAGGCCCTCCTGTCCTGCTCCACTAGGTGGCTGAAGTGACTGACAGC
1434
Not-treated
GtCtgtCTGTCCTGC --- FokI
FokI --- TGGCTGgctTtcgct
Dox-treated
0
1434

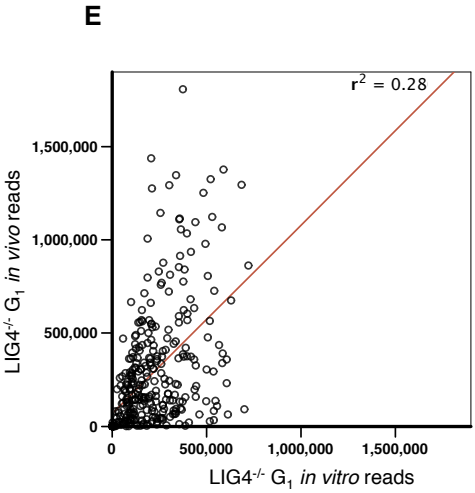
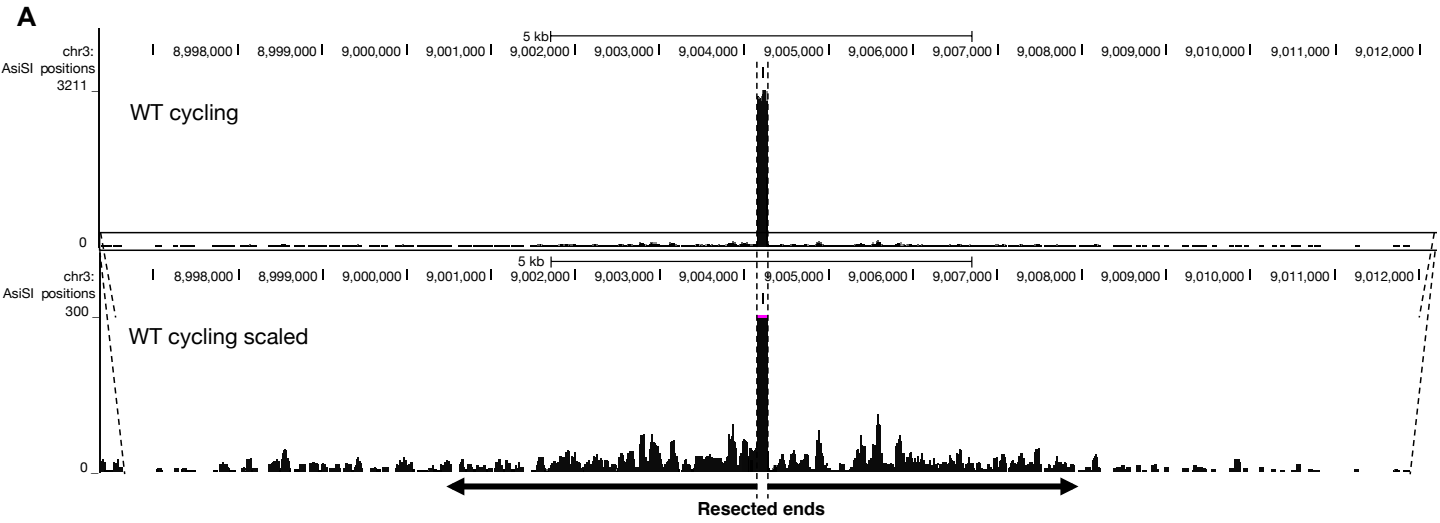
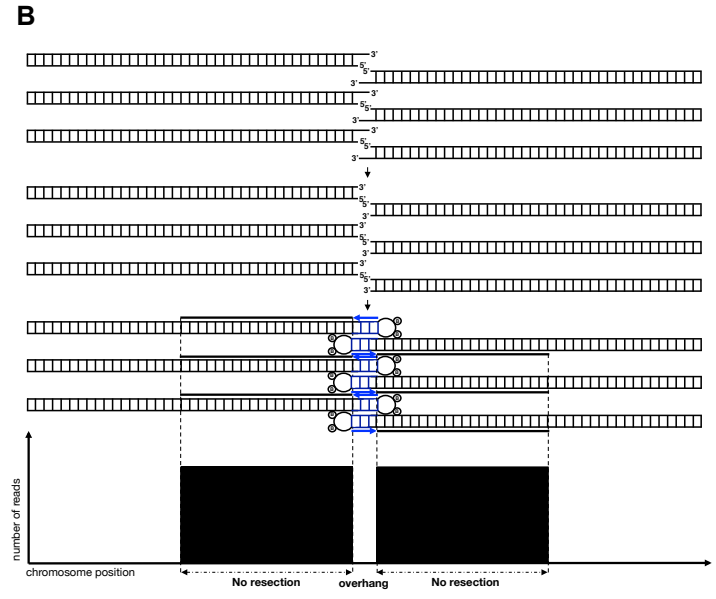

E
r2 = 0.28
1,500,000
1,000,000
500,000
0
LIG4-/- G1 in vivo reads
0
500,000
1,000,000
1,500,000
LIG4-/- G1 in vitro reads

Figure S3


A
chr3:
AsiSI positions
3211
WT cycling
0
5 kb
8,998,000
8,999,000
9,000,000
9,001,000
9,002,000
9,003,000
9,004,000
9,005,000
9,006,000
9,007,000
9,008,000
9,009,000
9,010,000
9,011,000
9,012,000
chr3:
AsiSI positions
300
WT cycling scaled
0
Resected ends


B
number of reads
chromosome position
No resection
overhang
No resection

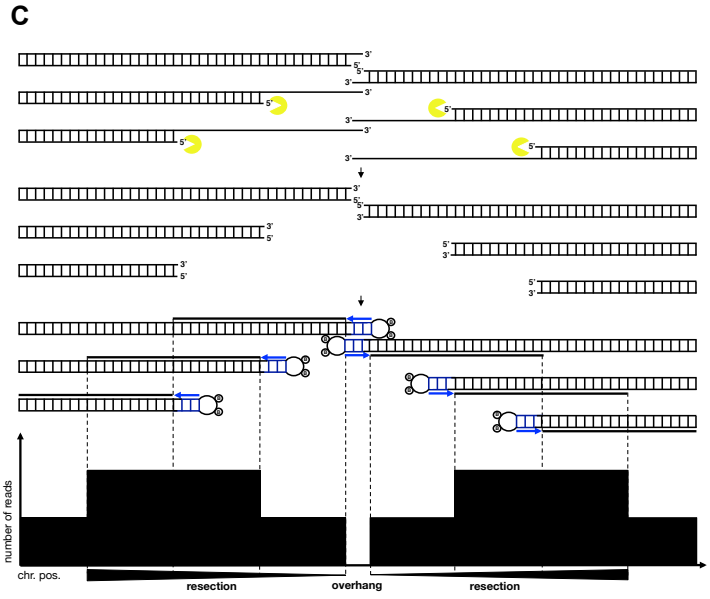

C
number of reads
chr. pos.
resection
overhang
resection

Figure S4

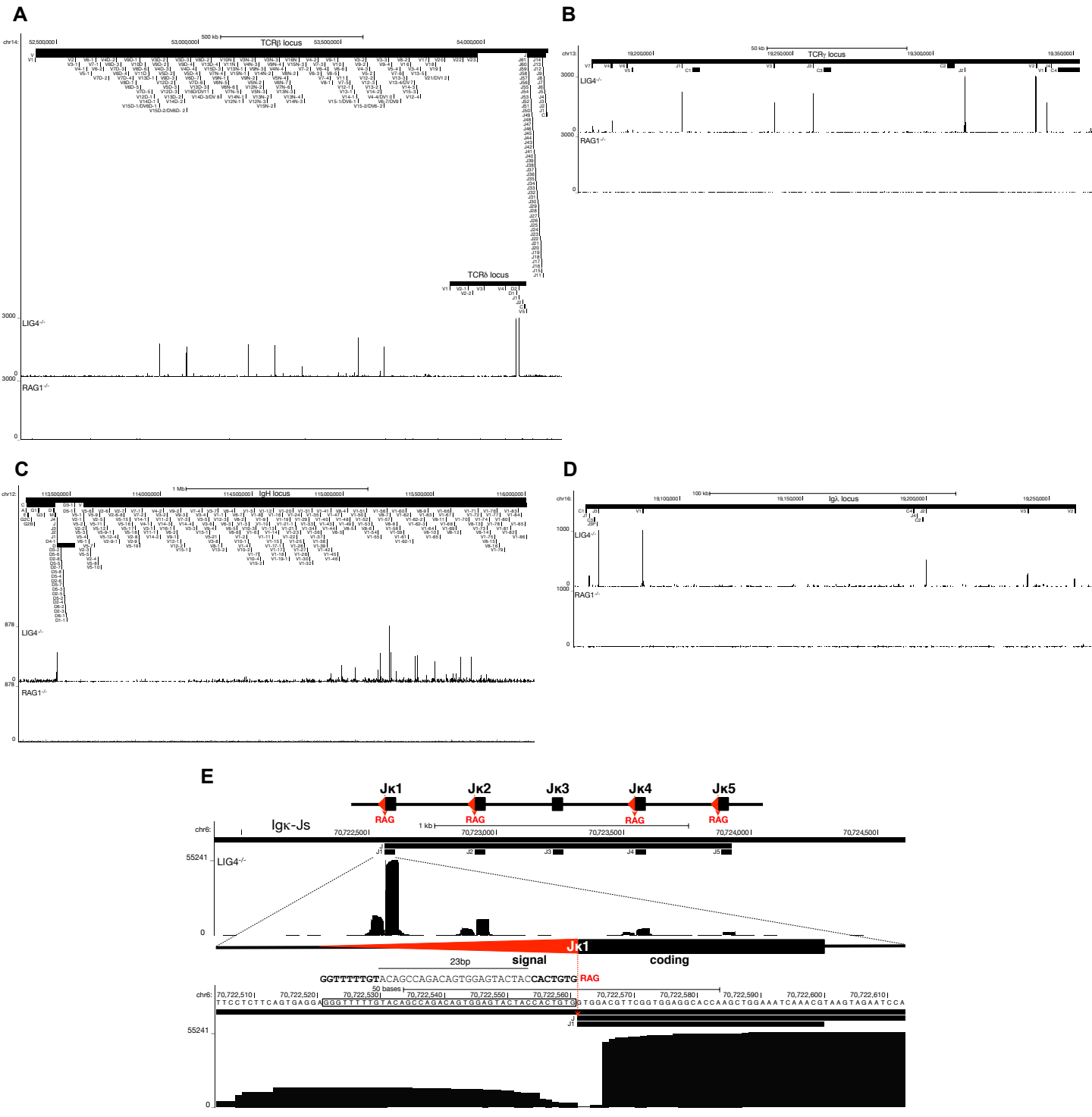
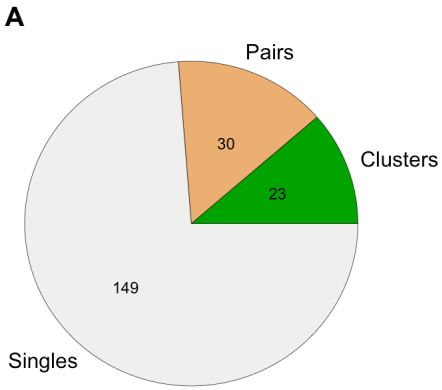

A
TCRβ locus
TCRδ locus
LIG4-/-
RAG1-/-
B
TCRγ locus
LIG4-/-
RAG1-/-
C
IgH locus
LIG4-/-
RAG1-/-
D
Igλ locus
LIG4-/-
RAG1-/-
E
Jκ1
Jκ2
Jκ3
Jκ4
Jκ5
RAG
Igκ-Js
LIG4-/-
23bp
signal
coding
GGTTTTTGTACAGCCAGACAGTGGAGTACTACCACTGTG
RAG
50 bases

Figure S5


A
Pairs
30
Clusters
23
149
Singles

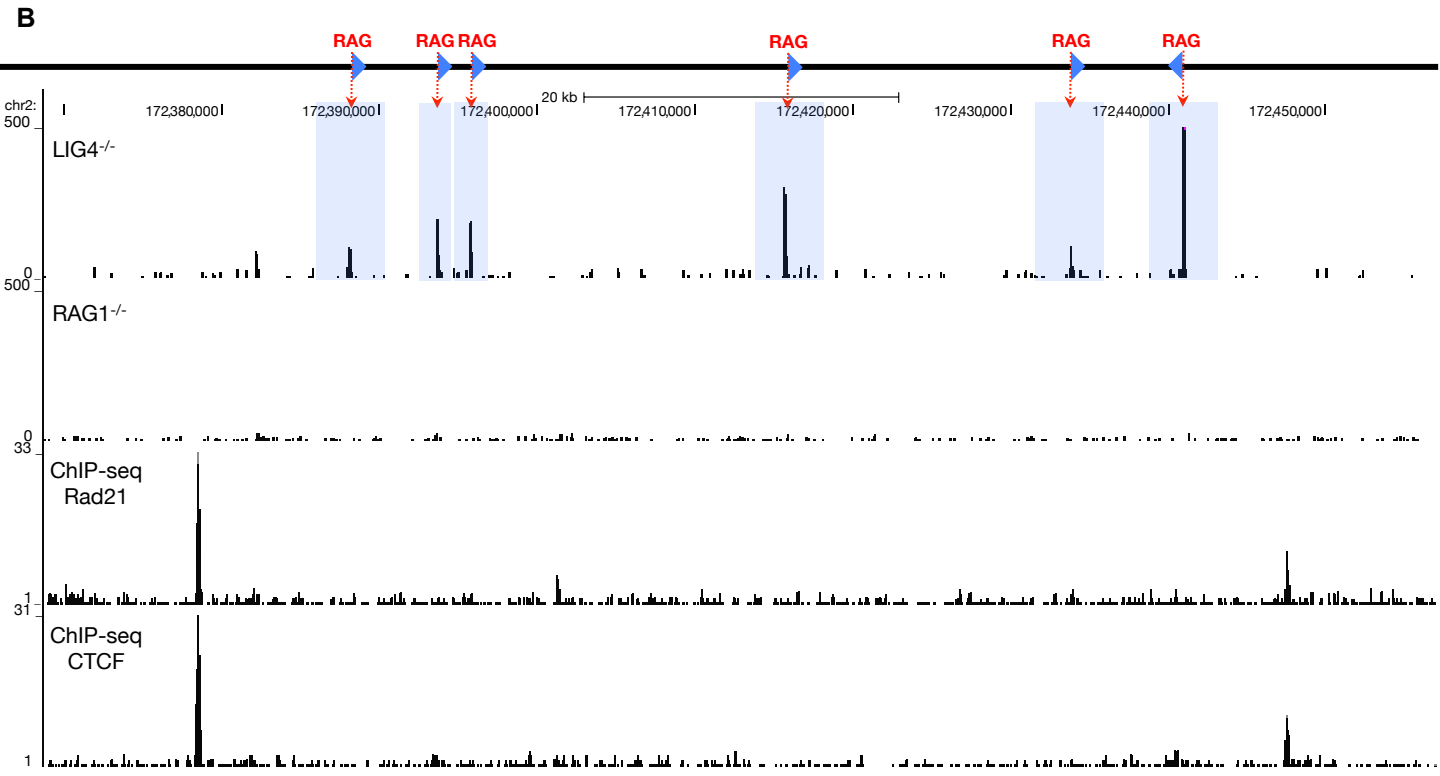

B
RAG
RAG
RAG
RAG
RAG
RAG
chr2:
172,380,000
172,390,000
172,400,000
20 kb
172,410,000
172,420,000
172,430,000
172,440,000
172,450,000
500
LIG4-/-
0
500
RAG1-/-
0
33
ChIP-seq
Rad21
1
31
ChIP-seq
CTCF
1

Figure S6

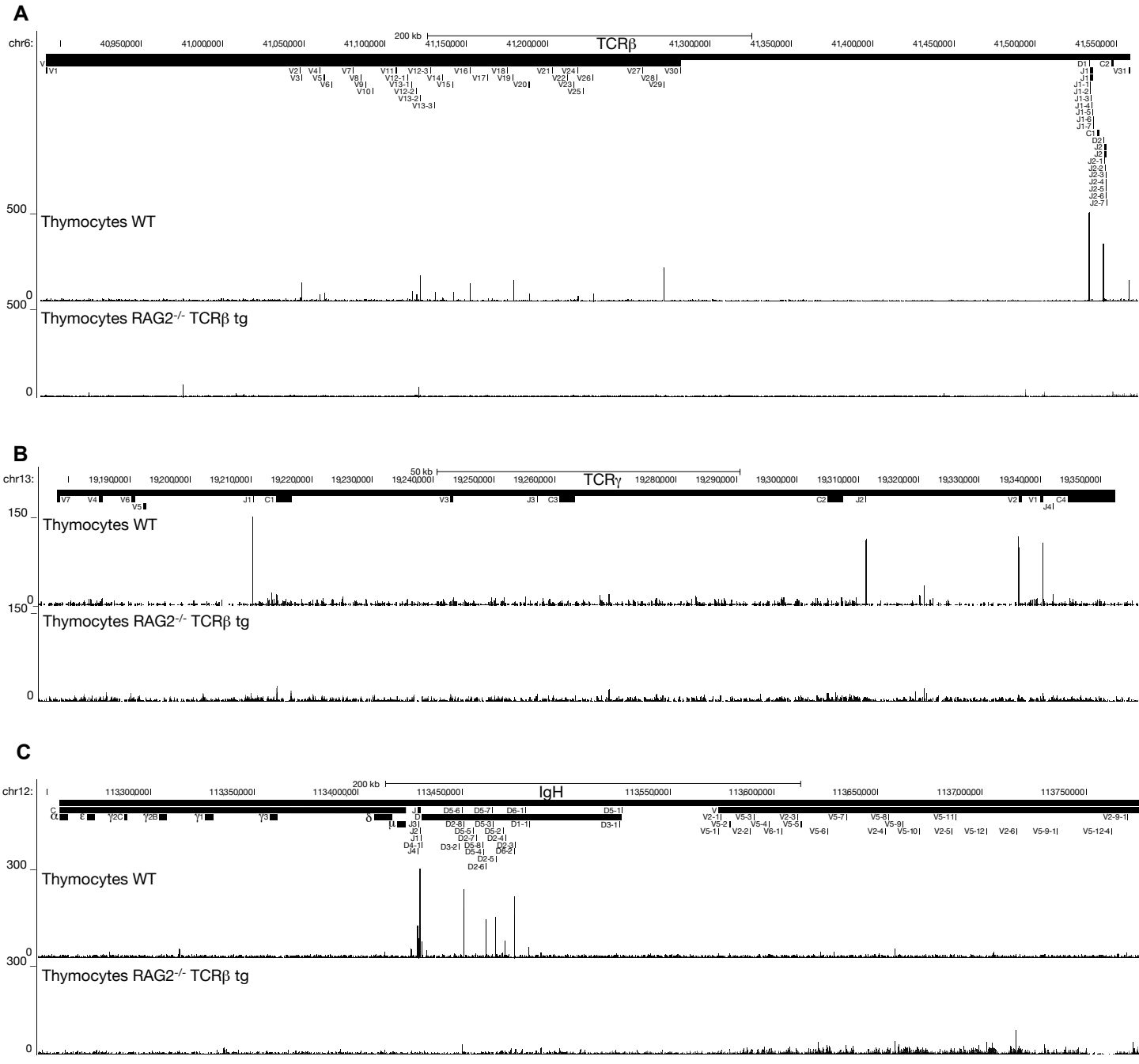
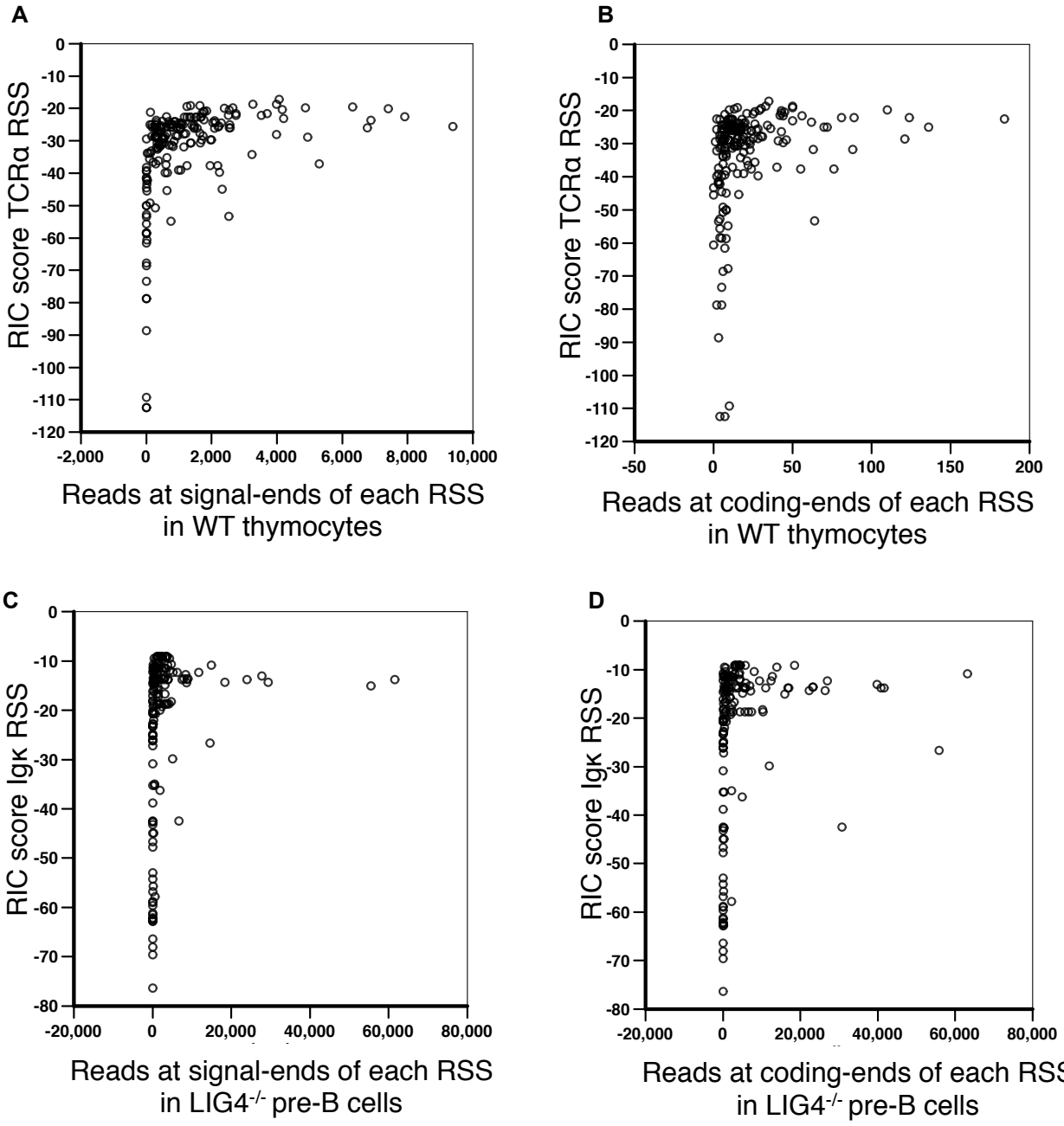

A
chr6:
200 kb
TCRβ
Thymocytes WT
Thymocytes RAG2-/- TCRβ tg
B
chr13:
50 kb
TCRγ
Thymocytes WT
Thymocytes RAG2-/- TCRβ tg
C
chr12:
200 kb
IgH
Thymocytes WT
Thymocytes RAG2-/- TCRβ tg

Figure S7


A
RIC score TCRα RSS
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
-110
-120
-2,000
0
2,000
4,000
6,000
8,000
10,000
Reads at signal-ends of each RSS
in WT thymocytes
B
RIC score TCRα RSS
-50
0
50
100
150
200
Reads at coding-ends of each RSS
in WT thymocytes
C
RIC score Igκ RSS
-80
-70
-60
-50
-40
-30
-20
-10
-20,000
0
20,000
40,000
60,000
80,000
Reads at signal-ends of each RSS
in LIG4-/- pre-B cells
D
RIC score Igκ RSS
Reads at coding-ends of each RSS
in LIG4-/- pre-B cells

SUPPLEMENTARY FIGURE AND TABLE LEGENDS

Figure S1. Breaks detected by γ-H2AX and END-seq in pre-B cells, related to Figure 1.

A. Schematic of DSB induction by *AsiSI*, ZFN and RAG endonuclease. DSBs can be produced either in G1 arrested cells (top panel, pre-B imatinib treatment for 48 hours) or in cycling cells (bottom panel). **B**. γ-H2AX (red) immunofluorescence (IF) in *LIG4*$^{-/-}$ pre-B cells after 48 hours imatinib treatment to produce RAG breaks **C.** γ-H2AX (red) immunofluorescence (IF) after 48 hours imatinib treatment together with DOX treatment during the last 24 hours and 4OHT treatment during the last 4 hours to simultaneously produce RAG-, ZFN- and *AsiSI*- breaks. **D.** Diagrammatic explanation of END-seq detection of an *AsiSI* break site. **E.** Reads from each end of the two-ended break have opposite divergent orientation. Top panel shows a track corresponding to the same *AsiSI* site shown in Figure 1B with all reads mapped without strand specificity. Below, the reads are split in those mapping to the plus strand (black) and to the minus strand (red). Bottom panel schematic explains how the reads are distributed into the negative strand (red) and positive strand (black).

Figure S2. *AsiSI* and ZFN on and off-target breaks, related to Figures 2 and 3.

A. END-seq reads on chromosome 8 for G1-arrested (top track) and cycling (bottom track) WT pre-B cells 4 hours after *AsiSI* induction. Predicted *AsiSI* sites are indicated above the first track by bars, and y-axes indicate number of reads. **B.** Pie chart showing different type of breaks detected in G1 arrested *LIG4*$^{-/-}$ pre-B cells (treated with imatinib, DOX and OHT) identified by peak calling. **C.** END-seq reads at the constant region of the

IgH locus in WT B cells stimulated for class switch recombination for 3 days (top data track) and input control (bottom data track). The positions of the constant, D and J gene segments are displayed on the top. Switching regions Sγ1 and Sμ are indicated. **D.** ZFN on-target (top panel) and a ZFN off-target (bottom panel) DSB detected in DOX treated G1 arrested *LIG4*$^{-/-}$ pre-B cells by END-seq. The sequence surrounding both ZFNs is illustrated, with perfect matches indicated in bold upper-case and mismatches in lower-case. **E.** Normalized read count at each broken *AsiSI* site in G1 arrested *LIG4* $^{-/-}$ pre-B cells induced for *AsiSI* cleavage *in vivo vs. in vitro* digestion with recombinant enzyme in the agarose plug.

Figure S3. DSB end-resection detected by END-seq, related to Figure 4.

A. Top panel shows the read coverage auto scaled to the highest peak at one *AsiSI* site detected in cycling WT pre-B cells. Bottom panel shows the magnification of the top panel, which reveals END-seq reads that are distal from the initial *AsiSI* site. **B.** Schematic representation of a non-resected end detected by END-seq. **C.** Schematic representation of a resected end detected by END-seq. 5'-3' nucleolytic degradation, represented in yellow, generates 3' overhangs, which are blunted at different positions, and give an irregular pattern of accumulated reads distal to the break site.

Figure S4. DSBs at antigen receptor loci in pre-B cells, related to Figure 5.

END-seq reads at TCRα **(A),** TCRγ **(B)** , IgH **(C)**, and Igλ **(D)** for *LIG4*$^{-/-}$ (top track) and *RAG1*$^{-}$ $^{/-}$ (botton track). Position of the V, D and J gene segments are indicated at the top. **E.**

Magnified view of RAG-induced DSBs mapped by END-seq at the Jκ locus. The position of the gene segments (black rectangles) and RSSs (red triangles) are indicated. Top panel shows reads at signal and coding ends for each gene segment. Bottom panel shows a magnification of the DSB at Jκ1. While RAG breaks at the signal-ends are precise and blunt, the irregular shape of the reads is caused by end-resection in the absence of LIG4.

Figure S5. RAG endonuclease on- and off-target activity, related to Figure 5.

A. Pie-chart classification of RAG off-targets. **B.** Example of RAG off-target DSBs in clusters. Tracks at the top show END-seq reads for *LIG4*$^{-/-}$ (first data track) and *RAG1*$^{-/-}$ (second data track), with off-target DSBs highlighted in blue. Blue triangles represent cryptic RSSs. Third and fourth data tracks show Rad21 and CTCF ChIP-seq respectively.

Figure S6. DSB repertoire in primary thymocytes, related to Figure 6.

END-seq reads at the TCRβ **(A)**, TCRγ **(B)** and IgH **(C)** loci for WT (top data track) and RAG2$^{-/-}$ TCRβ transgenic (bottom data track) thymocytes. The positions of all the V, D and J gene segments are displayed in the top.

Figure S7. Comparison between RIC store and DSB repertoire, related to Figures 6 and 7. **A-B.** Plots comparing RIC scores calculated for each RSS at the TCRα locus and total number of reads for signal **(A)** and coding **(B)** ends detected by END-seq in WT thymocytes. **C-D**. Same as (A, B) for Igκ locus in *LIG4*$^{-/-}$ pre-B cells.

Table S1. Comparison between END-seq and BLESS at *AsiS1* break sites, related to Figure 1.

Table S2. ZFN off-targets and *AsiS1* single nucleotide polymorphisms, related to Figures 1-3.

Table S3. RAG off-target DNA breaks associated with c-RSS detected in *LIG4*$^{-/-}$ pre B cells, related to Figure 5.

Table S4. Relationship between DSBs detected by END-seq, annotated RSSs and RIC scores, related to Figures 6 and 7.

Table S5. RAG off-target DNA breaks associated with c-RSS detected in WT and *ATM*$^{-/-}$ thymocytes, related to Figures 6 and 7.

Supplementary Experimental Procedures

Cell lines and Immunofluorescence.

Abelson-transformed pre-B cells (Bredemeyer et al., 2006; Lee et al., 2013) were retrovirally transduced with the tetracycline-inducible ER-AsiSI, pTRE3G-HA-ER-AsiSI as previously described (Santos et al., 2014). For RAG induction, pre-B cells were treated with 3 μM imatinib for 48h at 1 x 10^6 cells/ml. The ATM kinase inhibitor KU55933 (Tocris) was used at 15 μM. For induction of ZFN and ER-AsiSI, pre-B cells were treated with DOX at 1 μg/ml for 24h and with 4OHT at 1μM for 4 hours. Immunofluorescent staining with γ-H2AX (JBW301 Millipore, 1/10,000) antibody was performed in parallel in each pre-B cell experiment to verify a proper DSB induction.

Mice

C57BL/6 WT, *ATM*$^{-/-}$ (provided by A. Wynshaw-Boris), and *RAG2*$^{-/-}$ mice expressing a Vβ8 transgene (provided by K. Hathcock) between 4 and 18 weeks of age were used to prepare single cell suspensions from whole thymocytes and splenic B cells (purified using depletion with anti-CD43 beads). For class switch recombination experiments purified splenic B cells were cultured with LPS (25 μg/ml; Sigma), IL-4 (5 ng/ml; Sigma) and RP105 (Anti-Mouse CD180 (0.5 μg/ml; BD Pharmingen) for 3 days as previously described (Callen et al., 2007).

RAG off-targets and RIC scores

Following the mapping and aligning of sequence reads, a peak-calling algorithm (SICER) (Zang et al., 2009) was used to identify regions or peaks of significant enrichment around sites of DNA breaks. Each peak was then scanned for the RAG recognition motif - CAC/GTG, and the pattern of sequenced reads surrounding the motif was established. Given that reads linked with a specific RAG-targeted CAC/GTG are expected to be in proximity to that particular motif, a weighting or confidence index was developed as follows:by using a "*W*" base-pair (bp) window centered on each CAC/GTG motif, reads aligning within this window were determined and weighted by an index in which four possible combinations of reads relative to the break/motif site were considered. For each combination, we calculated a downstream and upstream read score from the weighted sums as shown below:

$$\text{Downstream read score} = \frac{\sum_{i=1}^{W} N_{i_{down}}^{+} * W_i}{\sum_{i=1}^{W} N_{i_{up}}^{+} * W_i}$$

$$\text{Upstream read score} = \frac{\sum_{i=1}^{W} N_{i_{up}}^{-} * W_i}{\sum_{i=1}^{W} N_{i_{down}}^{-} * W_i}$$

where, $N_{i_{down}}^{+}$ is the number of plus-strand reads at location '*i*' downstream of the DSB. Similarly, $N_{i_{up}}^{+}, N_{i_{up}}^{-}, N_{i_{down}}^{-}$ all denote numbers of reads at relevant locations. W_i indicates the weighting value used at location '*i*' away from the DSB.

RIC scores were calculated from annotated RSSs using the RSSsite (Merelli et al., 2010). For gene segments positions and RSS functionality for mouse C57BL/6J Immunoglobulin

and T cell receptor in GRCm38.p3 assembly were obtained from the IMGT Repertoire (Lefranc, 2001).

References

Bredemeyer, A.L., Sharma, G.G., Huang, C.Y., Helmink, B.A., Walker, L.M., Khor, K.C., Nuskey, B., Sullivan, K.E., Pandita, T.K., Bassing, C.H.*, et al.* (2006). ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. Nature *442*, 466-470.

Callen, E., Jankovic, M., Difilippantonio, S., Daniel, J.A., Chen, H.T., Celeste, A., Pellegrini, M., McBride, K., Wangsa, D., Bredemeyer, A.L*., et al.* (2007). ATM prevents the persistence and propagation of chromosome breaks in lymphocytes. Cell *130*, 63-75.

Lee, B.S., Gapud, E.J., Zhang, S., Dorsett, Y., Bredemeyer, A., George, R., Callen, E., Daniel, J.A., Osipovich, O., Oltz, E.M*., et al.* (2013). Functional intersection of ATM and DNA-dependent protein kinase catalytic subunit in coding end joining during V(D)J recombination. Molecular and cellular biology *33*, 3568-3579.

Lefranc, M.P. (2001). IMGT, the international ImMunoGeneTics database. Nucleic acids research *29*, 207-209.

Merelli, I., Guffanti, A., Fabbri, M., Cocito, A., Furia, L., Grazini, U., Bonnal, R.J., Milanesi, L., and McBlane, F. (2010). RSSsite: a reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. Nucleic acids research *38*, W262-267.

Santos, M., Faryabi, R.B., Ergen, A., Day, A., Malhowski, A., Canela, A., Onozawa, M., Lee, J.-E., Callen, E., Gutierrez-Martinez, P*., et al.* (2014). DNA-damage induced differentiation of leukemic cells as an anti-cancer barrier. Nature *514, 107-11*.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952-1958.