SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Generation of ATM-proficient and -deficient *c-Myc*-DJβ *v-Abl* pro-B Cell Lines

Cells from bone marrow of 3-weeks old *ATM*$^{-/-}$-*c-Myc*$^{DJβ/DJβ}$-*Eμ-Bcl-2* or *ATM*$^{+/-}$-*c-Myc*$^{DJβ/DJβ}$-*Eμ-Bcl-2* mice were transfected by pMSCV-v-Abl retrovirus as described previously (Bredemeyer et al., 2006). Transfected cells were cultured in RPMI medium with 15% (v/v) FBS for 2 months to form stable transformed *v-Abl* pro-B cell lines. ATM inhibitor Ku55933 was used at 10 μM to treat the WT DEL-SJ-containing cells.

Preparing HTGTS Libraries

Locus-specific primers for HTGTS were listed in Table S5. For preparing HTGTS libraries with DFL16.1 5' 12RSS bait at *IgH* locus, we used the same protocol described previously (Frock et al., 2015). For HTGTS libraries from *c-Myc*-DJβ cassette- or DEL-SJ-containing cells, we made some modifications to the original protocol as below, which showed the same pattern but with a higher efficiency. In the linear amplification-mediated PCR step, we used Phusion polymerase (Thermo) instead of Taq polymerase (Roche) to perform an 85-cycle PCR with 75-μM dNTPs. In the adapter ligation step, we used a new adapter (comprised of "GCGACTATAGGGCACGCGTGGNNNNNN-NH_2" and "/5'Phosphorylation/-CCACGCGTG CCCTATAGTCGC-NH_2") at the concentration of 2.5 μM instead. Restriction enzyme was used to suppress *bona fide* V(D)J recombination products within the paired *bona fide* RSSs-containing cassettes: *Pst*I for *c-Myc* E1 primer in *c-Myc*-DJβ cassette, *Sca*I for 12S primer in DEL-SJ, and *Mlu*I for 23S primer in DEL-SJ. No restriction enzyme blocking was performed for *IgH* and DEL-SJ GFP-primer HTGTS libraries. All the HTGTS libraries were sequenced by Miseq with paired ends each at a read length of 250 bp.

HTGTS Data Analyzing and Normalization

The pipeline used for running HTGTS was the same as described previously (Frock et al., 2015). Miseq sequencing reads were aligned to *mm9* genome, with modified genomes in some

cases depicted in details below. We further filtered the junctions by removing those with more than 30-bp un-aligned sequences between bait and prey sequences.

For HTGTS libraries from *c-Myc*-DJβ cassette- or DEL-SJ-containing cells, the sequencing reads were separately aligned to *mm9* genome or paired *bona fide* RSSs-containing cassettes. The former detects genome-wide translocations and the latter detects *bona fide* RSS-mediated rearrangements in the cassettes. Junctions in the paired *bona fide* RSSs-containing loop domains were displayed by IGV (Robinson et al., 2011) without adding the *bona fide* rearrangements. The junction diversity between RAG-initiated DSBs is very limited (Gellert, 2002), thus we took the duplicate junctions into account to reflect more accurate frequencies of translocations between bait *bona fide* RSSs and other RSSs. When junctions have the same bait length and prey length with the same length of inserted sequences if any, we defined them as duplicate junctions. The BM pro-B cells can proliferate to duplicate early translocation junctions, thus no duplicate junctions were contained for analysis in HTGTS libraries from these cells. For HTGTS libraries from the G1-arrested *v-Abl* pro-B cells, duplicate junctions with different prey length result from random sonication shearing on independent translocation junction-containing fragments. Therefore, these duplicate junctions were considered as independent biological events and were contained for analysis. For libraries from 12S or 23S primers in DEL-SJ-integrated cells, we prepared one library from pooled genomic DNA of 6 strains with independent integration locations for ATM-proficient and deficient *v-Abl* cells, respectively. All 12 DEL-SJ-containing loop domains showed very similar general patterns (see the text for details) providing robust controls for reproducibility of the main conclusions. For libraries with DFL16.1 5' 12RSS bait at *IgH* locus, to accurately map the junctions, Miseq reads were aligned to modified *mm9* genome (*mm9*_8653), in which the *IgH* sequence between DFL16.1 and J_H3 (*mm9* coordinates: chr12: 114667296 -114720403) was replaced by the DFL16.1-J_H3 join sequence (CTGCAGAGACAGTGACCAGAGTCCCTTGGCCCCAGTAAGCAAACCAGGGT AGCTACTACCGTAGTAATAAA) harbored in the $ATM^{-/-}$ *v-Abl* pro-B cells (8653). RAG on-target analysis included junctions within ±20 bp of the cleavage site of *IgH bona fide* RSSs and associated resection. Annotation of V_H RSSs was from Choi et al., 2013. Annotation of D_H and

J_H RSSs was obtained from IMGT/GENE-DB (Giudicelli et al., 2005). *IgH* junctions excluding on-targets were used for RAG off-target analysis. All duplicates were included for *IgH* junction analysis as on-target V(D)J recombination generates a vast majority of RSS joins that are perfectly fused and coding joins with very limited diversity as outlined above, thus, most independent V(D)J joins represented as duplicates in sequencing reads. We compared *IgH* off-target junctions containing all the duplicates with those containing only duplicates of diverse prey lengths and observed similar pattern between the two. To compare HTGTS libraries of ATM-deficient cells with and without IGCR1, the same amount (50 ug) of DNA were used for library preparation and junctions were further normalized to number of bait aligned reads that reflect the same amount of sequenced bait-containing alleles (Table S4). We also performed HTGTS from cycling cells before G1 arrest, which yielded less than 5% of junctions of G1-arrested cells, confirming that most of the junctions recovered from G1-arrested cells occurred *de novo* after G1 arrest. Note that the ChIP-seq data displayed for *IgH* locus was lifted-over to the *mm9*_8653 genome from the source (GSE40173).

Evaluation of On-target V(D)J Recombination Events in *c-Myc*-DJβ Cells

To estimate the frequency of on-target (within *c-Myc*-DJβ cassette) and off-target events in *c-Myc*-DJβ cells, we prepared a HTGTS library for each genotype of *c-Myc*-DJβ *v-Abl* and BM cells without suppression of on-target V(D)J recombination events. We found the on-target events within *c-Myc*-DJβ cassette in ATM-proficient and -deficient BM *c-Myc*-DJβ cells represent about 98% of all events in each genotype. The *c-Myc*-DJβ *v-Abl* cells contain a rearranged DJβ allele in both genotypes; therefore we excluded translocation reads with the same junctional nucleotides as the pre-arranged DJβ for analysis and calculated the percentages of on-targets as approximately 97 % for both *v-Abl* genotypes.

Identification of Genome-wide Cryptic RSSs

To identify genome-wide cryptic RSSs, we prepared libraries from DEL-SJ integrated in both chromosomes 1 and X, and 3 independent repeats for each. Translocation hotspots were

identified by MACS2 (Zhang et al., 2008) with extsize at 20 bp and false discovery rate (FDR) cut-off at 10^{-9}. Peak regions with HTGTS junctions occurring in at least three libraries were considered as recurrent hotspots. Hotspots within the antigen receptor loci and the two DEL-SJ-containing domains were excluded from analysis.

Consensus sequence logos for the heptamer of cryptic RSSs were generated by the weblogo online tool (http://weblogo.berkeley.edu). Although sequence logos were not generated for associated nonamers, additional 32-bp sequences associated with cryptic RSS heptamer containing potential nonamers were listed in Table S3.

Generation of Cas9/gRNA DSBs for HTGTS Libraries and Gene Deletions

ATM-deficient-DJβ, ATM-deificent, or ATM-deficient-DJβ-*Rag2*$^{-/-}$ *v-Abl* pro-B cells were treated with 3 μM STI-571 at a concentration of $5x10^5$ cells/ml for 30 hours followed by nucleofection of pX330-Cas9-cmyc plasmid (15μg) into $20x10^6$ G1-arrested cells (2 reactions; $10x10^6$ cells each reaction) using default program DN100 in SF solution (X unit of 4D-Nucleofector from Lonza). Transfected cells were cultured in RPMI medium with 15% (v/v) FBS plus 3 μM STI-571 for 3 more days before harvested for genomic DNA. HTGTS libraries were prepared and processed as described above; *Sph*I digestion was applied to suppress germ-line sequences during the preparation HTGTS libraries.

For the *Rag2* and IGCR1 deletion, a mixture of 2 μg upstream pX330-Cas9 plasmid and 2 μg downstream pX330-Cas9 plasmid was nucleofected into $3x10^6$ growing parental cells. Nucleofected cells were subjected to single-cell sub-cloning the next day. Sub-clones with correct genotypes were confirmed by both Sanger sequencing and Southern blot.

Mapping the Integration Sites of DEL-SJ

In order to map the integration sites of DEL-SJ, we designed a Cas9/gRNA cut site as bait in human CD4 in DEL-SJ (Table S5). To introduce breaks, 2 μg pX330-Cas9-hCD4 plasmid were transfected into $3x10^6$ DEL-SJ-containing *v-Abl* pro-B cells via nucleofection, and cells were harvested for DNA 3 days later. We generated HTGTS libraries and identified the DEL-SJ

integration sites by checking hybrid sequences containing the tail sequence of MMLV LTR and a piece of genomic sequence. Then the integration sites and orientations were confirmed by locus-specific PCR with a primer in human CD4 in DEL-SJ and the other primers in the genome close to the integration sites (Table S5).

Orientation of CBEs

To determine the orientation of CBEs, sequences of approximately 100 bp including core binding sites of CTCF were extracted from mouse *mm9* genome according to the ChIP-seq data (Lin et al., 2012). We then used ApE (http://biologylabs.utah.edu/jorgensen/wayned/ape/) to find the CBE motif with the least variations to the CBE consensus sequence 5'-CCACNAGGTGGCA G-3' (Rao et al., 2014). The CBE was defined as in "forward" orientation if the above consensus sequence reads from centromere to telomere, otherwise it's defined as in "reverse" orientation.

Distance to *bona fide* RSSs and CAC Motif

Loci information of CACs (and GTGs for reverse CACs) in the analyzed loop domains were derived from *mm9* genome by IGV (Robinson et al., 2011). Direct joining to CACs or the nucleotides adjacent to CACs was defined as 0.

Supplemental Table 2. Summary of 12 Loop Domains Revealed by DEL-SJ HTGTS Libraries, Related to Figure 3.

Genotype	Integrated chr	Integration site	Orientation*	Domain start	Domain end	Domain size
ATM$^{-/-}$	Chr1	34546111	-	34,275,801	34,904,745	629 kb
ATM$^{-/-}$	Chr4	94823238	-	94,675,379	95,583,722	908 kb
ATM$^{-/-}$	Chr5	63245462	-	61,747,765	64,203,530	2.46 Mb
WT	Chr6	31169213	-	30,999,620	31,576,644	577 kb
ATM$^{-/-}$	Chr8	97063319	-	97,053,683	97,227,492	174 kb
WT	Chr9	21133028	-	20,932,840	21,214,281	281 kb
ATM$^{-/-}$	Chr10	79579094	-	79,388,879	79,625,876	237 kb
WT	Chr12	78159092	+	77,908,979	80,012,123	2.10 Mb
WT	Chr12	114465716	+	114,461,087	114,667,734	207 kb
WT	Chr17	46173771	+	46,166,645	46,384,569	218 kb
WT	Chr19	29560017	-	28,851,550	29,895,679	1.04 Mb
ATM$^{-/-}$	ChrX	11242377	+	10,295,345	12,372,065	2.08 Mb

a. * "+" orientation is defined as 12RSS centromeric to 23RSS, or it's in "-" orientation.

Supplemental Table 4. Summary of HTGTS Libraries for *IgH* Locus, Related to Figure 6.

Genotype	Exp	Input DNA	Junctions with duplication*	*IgH* on-targets	*IgH* off-targets
ATM$^{-/-}$	repeat 1	50 ug	7,403	6,346	794
	repeat 2	50 ug	8,403	6,902	756
	repeat 3	50 ug	11,245	9,512	1,068
ATM$^{-/-}$-IGCR1Δ	repeat 1	50 ug	269,316	266,334	646
	repeat 2	50 ug	258,478	255,781	492
	repeat 3	50 ug	230,884	227,512	843

a. *Junctions are normalized to same amount of total bait aligned reads (606,723 reads per library). Bait aligned reads include all sequences containing bait primer, reflecting the same amount of bait alleles.

Supplemental Table 5. Oligos Used, Related to Experimental Procedures.

Name	Sequence	Purpose
Myc-E1-bio	/5bio/TGATGTTGGGCTAGCGCAG	HTGTS bio primers for DJβ cassette
Myc-E1-red	AGGGATGTGACCGATTCGTTG	HTGTS nested primers for DJβ cassette
12S-bio	/5bio/GACCTTACACAGTCCTGCTG	HTGTS bio primers for DEL-SJ 12RSS
12S-red	GTAGACGGCATCGCAGCTTG	HTGTS nested primers for DEL-SJ 12RSS
23S-bio	/5bio/CAAAAGACGGCAATATGGTGG	HTGTS bio primers for DEL-SJ 23RSS
23S-red	ACGCACACCGGCCTTATTCC	HTGTS nested primers for DEL-SJ 23RSS
GFP-bio	/5bio/GACAACCACTACCTGAGCAC	HTGTS bio primers for DEL-SJ GFP
GFP-red	CCAACGAGAAGCGCGATCAC	HTGTS nested primers for DEL-SJ GFP
Rag2-up	GAATAGGTCTTTTATCTGAA	Upstream Cas9 site for Rag2 deletion
Rag2-down	GAGCAATATACCTGAGTCTG	Downstream Cas9 site for Rag2 deletion
Myc-Cas9	GACGAGCGTCACTGATAGTA	For Cas9 site downstream of DJβ cassette
Myc-bio	/5bio/GCCTCGGCTCTTAGCAGACTG	HTGTS bio primers for myc-Cas9
Myc-red	CCTCTGAAGCCAAGGCCGATG	HTGTS nested primers for myc-Cas9
hCD4-Cas9	GCAGTGTCTGCTGAGTGACT	For Cas9 site at hCD4 in DEL-SJ
hCD4-bio	/5bio/TGATGAGAGCCACTCAGCTC	HTGTS bio primers for hCD4-Cas9
hCD4-red	CCTAAGCTGATGCTGAGCTTG	HTGTS nested primers for hCD4-Cas9
IS-hCD4	CAATGGCCCTGATTGTGCTG	In the DEL-SJ for verifying integration site
IS-Chr1	CCTATTTTCTCCGACTAGCC	For verifying DEL-SJ integration site
IS-Chr4	GAAGCACACATGTGGCTTGC	For verifying DEL-SJ integration site
IS-Chr5	CTGGTGCTTACACCATTAGAG	For verifying DEL-SJ integration site
IS-Chr6	GTGAGGTATGAAGCCATGACC	For verifying DEL-SJ integration site
IS-Chr8	TGTCGGGAGTTAGCTTCCAG	For verifying DEL-SJ integration site
IS-Chr9	CTGTCCTCACACTCAGGTTC	For verifying DEL-SJ integration site
IS-Chr10	CAAGTGGCGAACATGACCTAGC	For verifying DEL-SJ integration site
IS-Chr12	CTGCTACTATTGCTCCTCTG	For verifying DEL-SJ integration site
IS-Chr12	CTAAGACAGTAAGGTAGACACAC	For verifying DEL-SJ integration site
IS-Chr17	CTAGTTGGTTATGCGTATGAC	For verifying DEL-SJ integration site
IS-Chr19	GAAGTAGAACAGAAGGGAAAG	For verifying DEL-SJ integration site
IS-ChrX	CCGCACTCTTTGACACAGAC	For verifying DEL-SJ integration site
cRSS-bio	/5bio/GTTCATATTCAAAGTTTAGTCGCC	HTGTS bio primers for cryptic RSS at Chr1
cRSS-red	CTCTAGCCAAAAGCTGGCTTC	HTGTS nested primers for cryptic RSS at Chr1
IGCR1_up	GGAAAACTCTGTAGGACTAC	Upstream Cas9 site for IGCR1 deletion
IGCR1_down	TGGGACATGTAAACTGTAAC	Downstream Cas9 site for IGCR1 deletion
DFL16.1-bio	/5bio/CTTGCTCCCTAGGACCTTCC	HTGTS bio primer for DFL16.1 5' 12RSS
DFL16.1-red	ACTGAAACTCAACCGTGCTG	HTGTS nested primer for DFL16.1 5' 12RSS

a. DNA sequences of *c-Myc*-DJβ cassette and DEL-SJ are available on request.

SUPPLEMENTAL REFERENCES

Choi, N.M., Loguercio, S., Verma-Gaur, J., Degner, S.C., Torkamani, A., Su, A.I., Oltz, E.M., Artyomov, M., and Feeney, A.J. (2013). Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies. J. Immunol. *191*, 2393–2402.

Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. Annu. Rev. Biochem. *71*, 101–132.

Giudicelli, V., Chaume, D., and Lefranc, M.-P. (2005). IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res. *33*, D256–D261.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.