Supplemental Figures

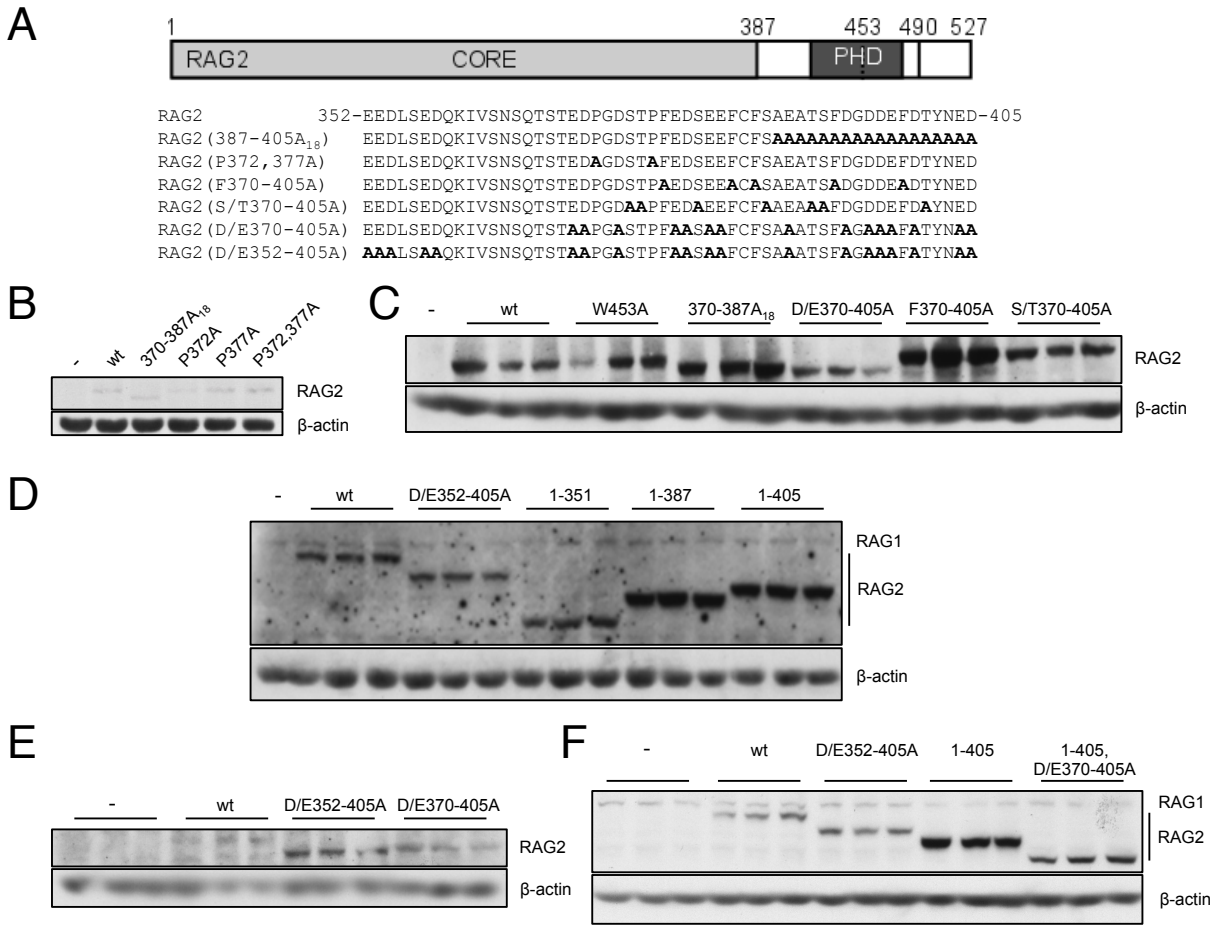

FIG S1 Related to Figure 1. (**A**) Above, diagram of RAG-2 with the canonical core (CORE) and plant homeodomain (PHD) indicated. Residues W453 and T490 are indicated by dotted and solid vertical lines, respectively. Below, amino acid sequences of the indicated RAG-2 mutants in the interval from residue 352 through 405, compared to the corresponding sequence of wild-type RAG-2. Mutated residues are indicated in bold type. (**B - F**) Upper, detection of RAG-2 variants assayed in Figs. 1B, 1C, 1D, 1E and 1F, respectively, by immunoblotting using an anti-myc antibody. Lower, immunodetection of actin on the same membranes as in the upper panels.

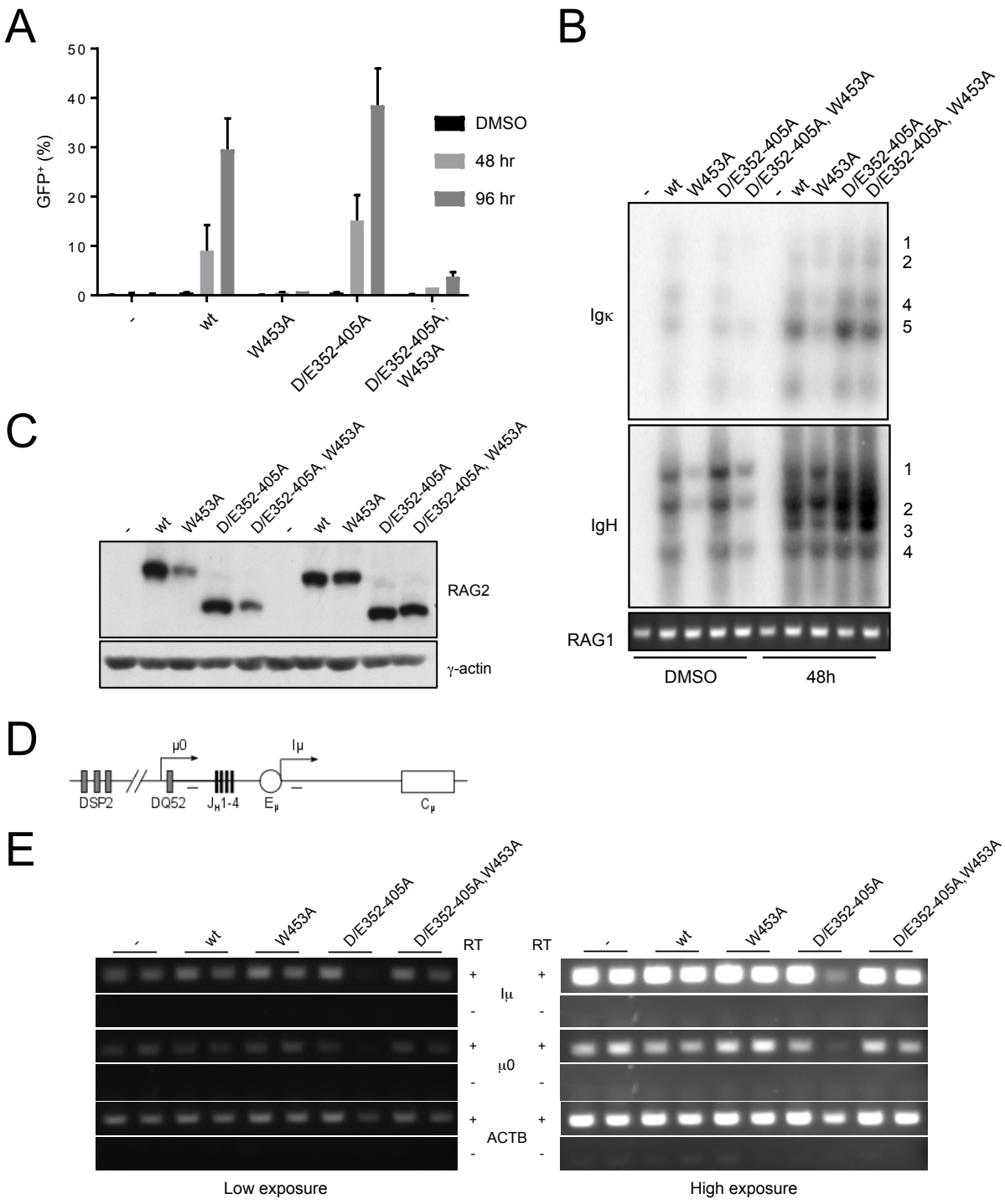

FIG S2 Related to Figure 3. (**A**) Assay of inversional recombination as in Fig. 4C for two additional independent infections of R2K3 cells. (**B**) Effects of mutations in the RAG-2 PHD finger and inhibitory domain on endogenous V(D)J recombination in R2K3 cells. Cells stably

expressing empty vector (-), wild-type RAG-2 (wt), or the indicated RAG-2 mutants were treated with DMSO or STI-571 for 48 hr. Top, recombination at the Ig κ locus was assayed as in Fig. 2. Middle, D_H-to-J_H recombination at the IgH locus, assayed as in Fig. 2. Bottom, a PCR amplicon from the RAG-1 locus as a control for input genomic DNA. (**C**) Immunodetection of myc-tagged RAG-2 variants in the transduced R2K3 cells assayed in (A). The two lanes corresponding to each variant contain protein from the two independent infections. Lower, immunodetection of actin on the same membrane as above. (**D**) A portion of the IgH locus is diagrammed with the origins of two germline transcripts indicated. The μ0 transcript is initiated at the DQ52 promoter. The PCR amplicon used to assay for this transcript is indicated as a horizontal line between DQ52 and the J_H cluster. Multiple Iμ transcripts originate from the Eμ enhancer 3' of the J_H cluster in the sense orientation. The PCR amplicon used to assay for these transcripts is marked by a horizontal line between Eμ and Cμ. (**E**) Assay for germline transcription at the IgH locus. RNA was isolated from two independent R2K3 transductants arrested with STI-571 for 48 hr. RNA samples were incubated in the presence (+) or absence (-) of reverse transcriptase and cDNA corresponding to Iμ (top), μ0 (middle) or actin (ACTB) (bottom) was detected by PCR. Left and right, short and long exposures, respectively, of the same gel, stained with ethidium bromide.

A

```
Plasmid      TGTTTTTGTTCCAGTCTGTAGCACTGTG CAGGTCTC CTGAACCTG CACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC

Wild-type cells
wt (4/4)     TGTTTTTGTTCCAGTCTGTAGCACTGTG                    CACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
352N (5/5)   TGTTTTTGTTCCAGTCTGTAGCACTGTG                    CACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
DM (5/6)     TGTTTTTGTTCCAGTCTGTAGCACTGTG                    CACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
DM           TGTTTTTGTTCCAGTCTGTAGCACTGTG                    --CAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC

DNA-PKcs-deficient cells
352N (5/10)  TGTTTTTGTTCCAGTCTGTAGCACTGTG                    CACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTGTAGCAC----                    ---AGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTGTAGCAC----                    ---AGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTGTAGCA-----                    ----GTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTGTAGCACTGT-                    ----------------------CTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTG---------- A                  ---------AGTACTCCACTGTCTGGCTGTACAAAAACC
DM (2/5)     TGTTTTTGTTCCAGTCTGTAGCACTGTG                    CACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
DM           TGTTTTTGTTCCAGTCTGTAGCAC----                    ---AGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
DM           TGTTTTTGTTCCAGTCTGTAGCACTGT-                    ------------------------------ACAAAAACC
DM           -125bp (entire RSS deleted)

XRCC4-deficient cells
352N         TGTTTTTGTTCCAGTCTGTAGCACTGTG                    ----GTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTGTAGCACTGTG                    -------GTAGTACTCCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTGTAGCAC----                    ---AGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCNGTCTGTAG-------                    -----------TACTCCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTGTAGC------                    ------------CTCCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCT-----------                    ---------------CCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTGTAGCACT---                    ---------------CCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTGTAGCAC---- T                  ----GTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTGTAGCACTGTG CTT                ---AGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
352N         TGTTTTTGTTCCAGTCTGTAGCACTG--         CTGAACCTG CACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
DM           TGTTTTTGTTCC----------------                    --CAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
DM           TGTTTTTGTTCC----------------                    --CAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC
DM           TGTTTTTGTTCCAGTCTGTAGCACTGT-                    ---------AGTACTCCACTGTCTGGCTGTACAAAAACC
DM           TGTTTTTGTTCCAGT-------                          ------------------------------AAAAACC
DM           -12bp                                          ------------------CTGTCTGGCTGTACAAAAACC
DM           -12bp                                          --------------------GTCTGGCTGTACAAAAACC
DM           TG-------------------------- (24bp untemplated)                                      -1
DM           -300+bp (entire RSS deleted)
DM           -300+bp (entire RSS deleted)
```

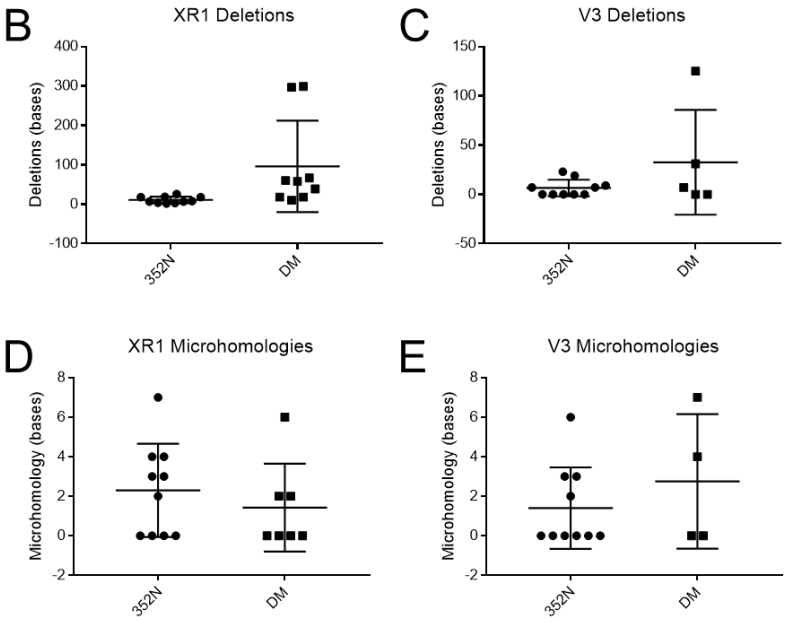

FIG S3 Related to Figure 4. (**A**) Signal joints recovered from wild-type CHO cells, XRCC4-deficient cells or DNA-PK_{CS}-deficient cells expressing wild-type RAG-2 (wt), RAG-2(D/E352-405A) (352N) or RAG2(D/E352-405A, W453A) (DM). The unrecombined RSSs of pJH200 and a portion of the flanking sequences are shown at top (Plasmid). Heptamer elements are shown in blue. Each section lists the signal joint sequences obtained from one of the three cell lines assayed, expressing the RAG-2 variant indicated at left. The number of precise signal joints relative to the total number of unique joints is given in parentheses. Microhomologies are designated by bold type and deletions by dashed lines. A templated insertion is shown in green. Only unique sequences from each transformation are included. (**B**) Lengths of deletions at signal joints from XRCC4-deficient cells. 352N, RAG2(D/E352-405A); DM, RAG2(D/E352-405A, W453A) Mean ± SD are indicated; means are significantly different ($p < 0.005$) by Kruskal-Wallis test. (**C**) Lengths of deletions at signal joints from DNA-PK_{CS}-deficient cells, displayed as in (B). Means are not significantly different ($p > 0.05$). (**D**) Lengths of microhomologies at signal joints from XRCC4-deficient cells, as in (B). Means are not significantly different ($p > 0.05$). (E) Lengths of microhomologies at signal joints from DNA-PK_{CS}-deficient cells, as in (B). ns, $p > 0.05$.

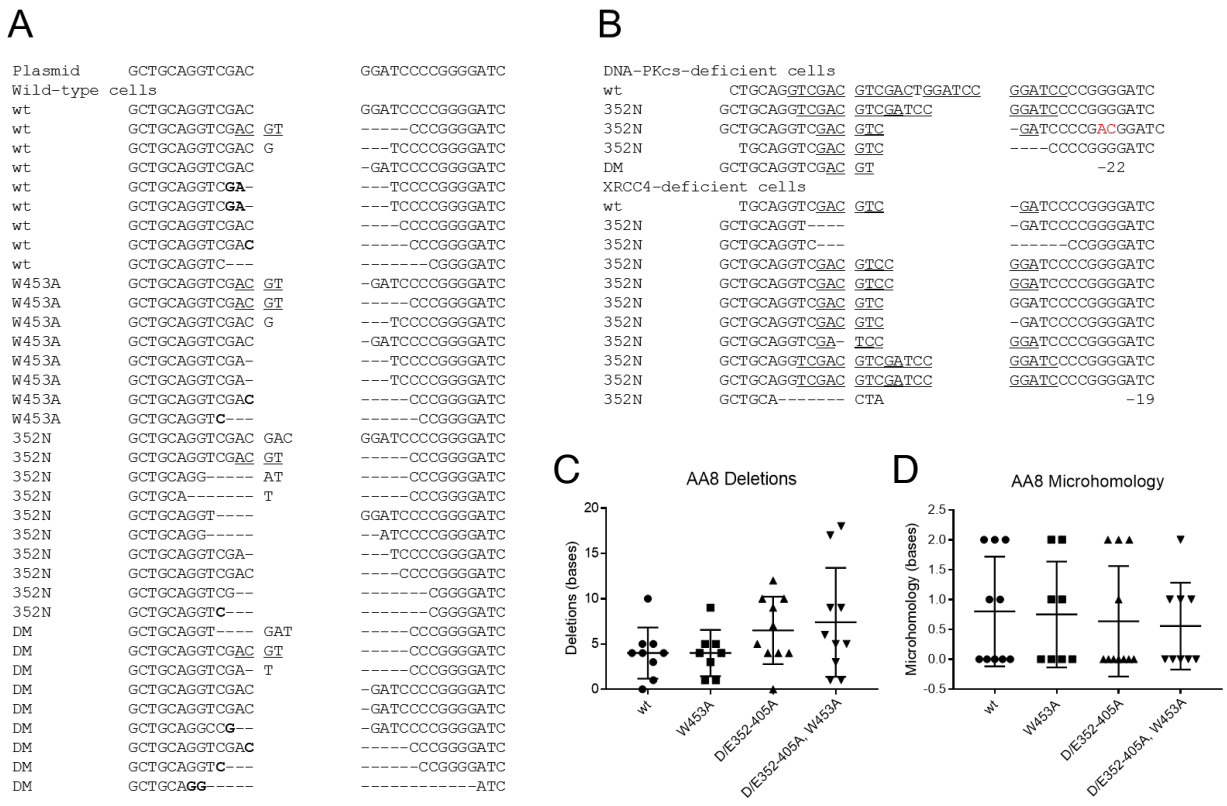

FIG S4 Related to Figure 4. Coding joints recovered from wild-type CHO cells (**A**) or NHEJ-deficient cells (**B**) expressing wild-type RAG-2 (wt), RAG-2(D/E352-405A) (352N) or RAG2(D/E352-405A, W453A) (DM). The unrecombined coding sequences flanking the RSSs of pJH200 are shown at top (Plasmid). Microhomologies are designated by bold type and deletions by dashed lines. P insertions are underlined and potential microhomologies within the P insertions are marked by thick underlining. A mutation from the plasmid sequence is indicated in red type. Six complex rearrangements were ambiguous and excluded. Only unique sequences from each transformation are included. (**C**) Lengths of deletions at coding joints from wild-type cells. Mean ± SD are indicated; no significant differences were observed ($p > 0.05$, Kruskal-Wallis). (**D**) Lengths of microhomologies at coding joints from wild-type cells. Means were not significantly different ($p > 0.05$, Kruskal-Wallis).

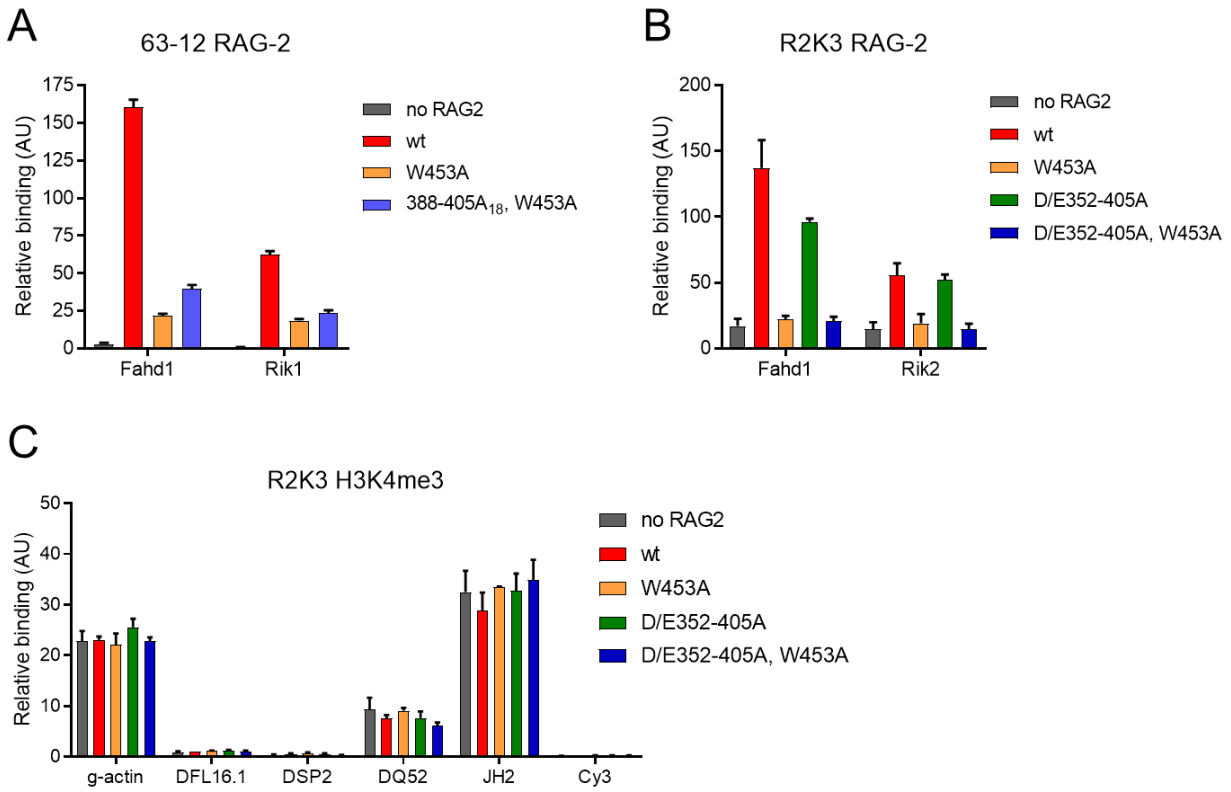

FIG S5 The PHD of RAG-2 mediates preferential binding to sites of H3K4 trimethylation. Assay for binding of RAG-2 to transcribed, non-immunoglobulin loci in 63-12 cells (**A**) and R2K3 cells (**B**) by ChIP. Mean ± SD are indicated; n=2. (**C**) Distribution of H3K4me3 at the IgH locus in R2K3 cells, assayed by ChIP. Enrichment of H3K4me3 over each of the genomic regions defined below was assayed by qPCR as described (Ji et al. 2010). Mean ± SD are indicated; n=2.

Supplemental materials and methods

Expression constructs. RAG-1 and RAG-2, tagged at the N terminus with maltose-binding protein (MBP) tag and at the C-terminus with a c-myc epitope and a polyhistidine sequence, were expressed in the vectors pcDNA1 and pcDNA3.1 respectively. Methods and oligonucleotide sequences used to generate RAG-2 variants are given in Supplemental Table 1. For GeneArt String (Thermo Fisher) constructs, endogenous PasI and PflMI restriction sites were used. Truncation mutations were introduced by PCR using the forward primer indicated as RAG-2 truncation F and appropriate reverse primers followed by cleavage with BamHI and EcoRI.

Supplemental Table 1. Cloning methods and oligonucleotide reagents for generation of RAG2 variants.

Mutation	Cloning Method	Primers 5' to 3'
RAG2(P372A)	Site Directed Mutageneis	F: TGGAGTCACCAGCATCTTCTGTTGATGTCTGACT R: AGTCAGACATCAACAGAAGATGCTGGTGACTCCA
RAG2(P377A)	Site Directed Mutageneis	F: TGAGTCTTCAAAGGCAGTGGAGTCACCAGGATC R: GATCCTGGTGACTCCACTGCCTTTGAAGACTCA
RAG2(P372,377A)	Site Directed Mutageneis	Sequential use of primers for P372A and P377A
RAG2(370-387A_{18})	Divergent PCR	F: GCCGCTGCAGCTGCTGCTGCTGCTGCTGCT GCTGCTGAAGCAACCAGTTTTG R: GCCGCTGCAGCGGCGGCGGCGGCGGCGGCGG CTGTTGATGTCTGACTGTTG
RAG2(D/E370-405A)	GeneArt Strings	GGCGCCCCCTGGGTACCCCAGCAGTGAATTGCACAGTCTTG CCAGGAGGAATCTCTGTCTCCAGTGCAATCCTCACTCAAAC AAACAATGATGAATTTGTTATTGTGGGTGGTTATCAGCTGG AAAATCAGAAAAGGATGGTCTGCAGCCTTGTCTCTCTAGGG GACAACACGATTGAAATCAGTGAGATGGAGACTCCTGACTG GACCTCAGATATTAAGCATAGCAAAATATGGTTTGGAAGCA

		ACATGGGAAACGGGACTATTTTCCTTGGCATACCAGGAGAC AATAAGCAGGCTATGTCAGAAGCATTCTATTTCTATACTTT GAGATGCTCTGAAGAGGATTTGAGTGAAGATCAGAAAATTG TCTCCAACAGTCAGACATCAACAGCAGCTCCTGGTGCCTCC ACTCCCTTTGCAGCCTCAGCCGCATTTTGTTTCAGTGCTGC AGCAACCAGTTTTGCTGGTGCCGCTGCATTTGCCACCTACA ATGCAGCTGATGAAGATGACGAGTCTGTAACCGGCTACTGG ATAACATGTTGCCCTACTTGTGATGTTGACATCAATACCTG GGTTCCGTTCTATTCAACGGAGCTCAATAAACCCGCCATGA TCTATTGTTCTCATGGGGATGGGCACTGGGTACATGCCCAG TGCATGGGGCGCC
RAG2(P370-405A)	GeneArt Strings	CCGCGGCCCTGGGTACCCCAGCAGTGAATTGCACAGTCTTG CCAGGAGGAATCTCTGTCTCCAGTGCAATCCTCACTCAAAC AAACAATGATGAATTTGTTATTGTGGGTGGTTATCAGCTGG AAAATCAGAAAAGGATGGTCTGCAGCCTTGTCTCTCTAGGG GACAACACGATTGAAATCAGTGAGATGGAGACTCCTGACTG GACCTCAGATATTAAGCATAGCAAAATATGGTTTGGAAGCA ACATGGGAAACGGGACTATTTTCCTTGGCATACCAGGAGAC AATAAGCAGGCTATGTCAGAAGCATTCTATTTCTATACTTT GAGATGCTCTGAAGAGGATTTGAGTGAAGATCAGAAAATTG TCTCCAACAGTCAGACATCAACAGAAGATCCTGGTGACTCC ACTCCCGCTGAAGACTCAGAGGAAGCTTGTGCCAGTGCTGA AGCAACCAGTGCTGATGGTGACGATGAAGCTGACACCTACA ATGAAGATGATGAAGATGACGAGTCTGTAACCGGCTACTGG ATAACATGTTGCCCTACTTGTGATGTTGACATCAATACCTG GGTTCCGTTCTATTCAACGGAGCTCAATAAACCCGCCATGA TCTATTGTTCTCATGGGGATGGGCACTGGGTACATGCCCAG TGCATGGGGCGCC
RAG2(S/T370-450A)	GeneArt Strings	CCGCGGCCCTGGGTACCCCAGCAGTGAATTGCACAGTCTTG CCAGGAGGAATCTCTGTCTCCAGTGCAATCCTCACTCAAAC AAACAATGATGAATTTGTTATTGTGGGTGGTTATCAGCTGG AAAATCAGAAAAGGATGGTCTGCAGCCTTGTCTCTCTAGGG GACAACACGATTGAAATCAGTGAGATGGAGACTCCTGACTG GACCTCAGATATTAAGCATAGCAAAATATGGTTTGGAAGCA ACATGGGAAACGGGACTATTTTCCTTGGCATACCAGGAGAC AATAAGCAGGCTATGTCAGAAGCATTCTATTTCTATACTTT GAGATGCTCTGAAGAGGATTTGAGTGAAGATCAGAAAATTG TCTCCAACAGTCAGACATCAACAGAAGATCCTGGTGACGCC GCTCCCTTTGAAGACGCAGAGGAATTTTGTTTCGCTGCTGA AGCAGCCGCTTTTGATGGTGACGATGAATTTGACGCCTACA ATGAAGATGATGAAGATGACGAGTCTGTAACCGGCTACTGG ATAACATGTTGCCCTACTTGTGATGTTGACATCAATACCTG GGTTCCGTTCTATTCAACGGAGCTCAATAAACCCGCCATGA TCTATTGTTCTCATGGGGATGGGCACTGGGTACATGCCCAG TGCATGGCCGCGG
RAG2(D/E352-405A)	GeneArt Strings	GGCGCCCCCTGGGTACCCCAGCAGTGAATTGCACAGTCTTG CCAGGAGGAATCTCTGTCTCCAGTGCAATCCTCACTCAAAC

		AAACAATGATGAATTTGTTATTGTGGGTGGTTATCAGCTGG AAAATCAGAAAAGGATGGTCTGCAGCCTTGTCTCTCTAGGG GACAACACGATTGAAATCAGTGAGATGGAGACTCCTGACTG GACCTCAGATATTAAGCATAGCAAAATATGGTTTGGAAGCA ACATGGGAAACGGGACTATTTTCCTTGGCATACCAGGAGAC AATAAGCAGGCTATGTCAGAAGCATTCTATTTCTATACTTT GAGATGCTCTGCGGCCGCTTTGAGTGCAGCTCAGAAAATTG TCTCCAACAGTCAGACATCAACAGCAGCTCCTGGTGCCTCC ACTCCCTTTGCAGCCTCAGCCGCATTTTGTTTCAGTGCTGC CGCAACCAGTTTTGCTGGTGCCGCTGCATTTGCCACCTACA ATGCAGCTGATGAAGATGACGAGTCTGTAACCGGCTACTGG ATAACATGTTGCCCTACTTGTGATGTTGACATCAATACCTG GGTTCCGTTCTATTCAACGGAGCTCAATAAACCCGCCATGA TCTATTGTTCTCATGGGGATGGGCACTGGGTACATGCCCAG TGCATGGGGCGCC
pCLIP2A	PCR	F: GCGATCGATACCATGTCCCTGCAGATGGTAAC R: GCGCTCGAGCGGCCGCTTATAACGCGCGC
RAG2 truncation F	PCR	F: GCTCGGATCCCGGGTACC
RAG2(1-351) R	PCR	R: GAATTCAGAGCATCTCAAAGTATAGAAATA
RAG2(1-387) R	PCR	R: GAATTCACTGAAACAAAATTCCTCTGAG
RAG2(1-405) R	PCR	R: GAATTCATCTTCATTGTAGGTGTCAAATTC
RAG2(1-405, D/E352-405A) R	PCR	R: GAATTCAGCTGCATTGTAGGTGGCAAA

Assays for recombination of endogenous immunoglobulin gene segments. The retroviral vector pCLIP2A (Pomerantz et al. 2002) was programmed to coexpress puromycin N-acetyl transferase and RAG-2 variants. The RAG-2 cassettes were amplified from the corresponding pcDNA3.1 subclones by PCR using primers indicated in Table 1. Viral particles were generated by cotransfection of pCLIP2A constructs and pCL-Eco into 293T cells and concentrated by centrifugation. The B progenitor cell lines 63-12 and R2K3 were infected by spin inoculation in the presence of 10 μg/mL polybrene. Cells were maintained under selection with 1 μg/mL puromycin for 21 - 25 days, at which time they were assayed for recombination. Recombination

was induced in R2K3 cells (10 ml at 10^6 cells/ml) by addition of STI-571 (3 mM stock in DMSO) to a concentration of 3 μM for 48 or 96 hr.

Genomic DNA was isolated from approximately 10^6 cells (Qiagen Dneasy) and D_H-to-J_H or Vκ-to-Jκ rearrangements were detected by PCR. Rearrangements were amplified from 100 ng genomic DNA template by PCR using primers and annealing temperatures indicated in Table 2. RAG-1 was amplified for purposes of normalization from 30 ng genomic DNA template. Reaction products were fractionated by electrophoresis through 1% agarose, transferred to Hybond N+ membranes (GE Healthcare) and detected by hybridization to radiolabeled probes defined in Supplemental Table 2. For quantitation, background was subtracted using a rolling ball algorithm and band area was determined by ImageJ.

Assay for coupled DNA cleavage

Coupled cleavage assays were performed using the duplex oligonucleotide substrates HL44/45 (12-RSS) and HL46/47 (23-RSS) (1) as described (2). In brief, equivalent amounts of active wild-type or mutant RAG tetramer (1 nM) were incubated with 5 nM ^{32}P-labeled HL44/45, 5 nM unlabeled HL46/47 and varying amounts of H3K4me3 or H3K4me0 peptide in 10 µL of a reaction buffer containing 25 mM 3-(4-Mopholino)propane sulfonic acid-KOH pH 7.0, 30mM KCl, 30 mM potassium glutamate, 5 mM $CaCl_2$, 0.5 mM DTT, 0.1 mg/mL BSA and 1% glycerol. After 20 min at 37°C, $MgCl_2$ was added to 5 mM and incubation was continued for 1 hr. Reactions were stopped by addition of loading buffer (90% formamide, 0.2% bromophenol blue) and incubation at 90°C for 2 min. Products were fractionated on a 15% polyacrylamide-urea gel, detected by phosphorimaging and quantified using ImageQuantNL.

Supplemental Table 2. Primers and probes used to assay rearrangement in B-cell progenitors.

Primer	Anneal/Hybridization (°C)	Sequence 5' to 3'
DSP2 F	65 Anneal	ATGGCCCCTGACACTCTGCACTGCT
DFL16.1 F	65 Anneal	ACACCTGCAAAACCAGAGACCATA
Jh4 R	65 Anneal	AAAGACCTGGAGAGGCCATTCTTACC
Jh probe	61 Hybridization	CTTACCTGAGGAGACGGTGAC
IgKv F	65 Anneal	GSTTCASWGGCAGTGGRTCTGG
Mar35 R	65 Anneal	AACACTGGATAAAGCAGTTTATGCCCTTTC
IgK probe	58 Hybridization	GCTCATTATCAGTTGACGTGGC
RAG1 F	Anneal	GCATCTATTCTGTAGGATCTGC
RAG1 R	Anneal	AAACAATGTCAAGCAGACAGCC
Imu F	60 Anneal	AATACCCGAAGCATTTACAGTGACT
Imu R	60 Anneal	AAGATTTGTGAAGCCGTTTTGACCA
Mu0 F	60 Anneal	GTGCAGGTCCCTCTCTTGTT
Mu0 R	60 Anneal	GACATTGCATCCACCCTTCT
ACTB F	61 Anneal	TGACGTGGACATCCGCAAAG
ACTB R	61 Anneal	CTGGAAGGTGGACAGCGAGG

Supplemental Table 3. Primers used in ChIP analyses.

Primer	Sequence 5'-3'	Source
γ-actin FP	GACACCCAACCCCGTGACG	(Subrahmanyam et al. 2012)
γ-actin RP	GCGGCCATCACATCCCAG	
Cγ3 FP	TGGACAAACAGAAGTAGACATGGGTC	
Cγ3RP	GGGGTTTAGAGGAGAGAAGGCAC	
DSP2s FP	TGTTACCTTACTTGGCAGGGATTT	
DSP2s RP	TGGGTTTTTGTTGCTGGATATATC	
DFL16.1 FP	CAAAGCAGCCACCATCCAG	(Chakraborty et al. 2009)
DFL16.1 RP	GCAGCACGGTTGAGTTTCAG	
DQ52 FP	CCCTGTGGTCTCTGACTGGTG	
DQ52 RP	GATTTCTCAAGCCTCTCTACTTCCTC	
JH2 FP	TACTTTGACTACTGGGGC	

JH2 RP	CCCTAGTCCTTCATGACC	
Jκ1 FP	TTGTACAGCCAGACAGTGGAG	(Ji et al. 2010)
Jκ1 RP	GCCACAGACATAGACAACGG	
Jκ2 FP	CAGATTCTGGCACTCTCCAA	
Jκ2 RP	ACTGAGCATGGTCTGAGCAC	
Jκ4 FP	AGTGTGAAAGCTGAGCGAAA	
Jκ4 RP	CACAGTGAGGACTATGACATGC	
Cκ FP	GTGGAAGATTGATGGCAGTG	
Cκ RP	GCTCATGCTGTAGGTGCTGT	
Fahd1FP	AGAGACCTTTCGCTGACCTC	
Fahd1RP	GGTCATGTGACCACCGACT	
Rik FP	GTTTCCACCGGAAGTGCT	
Rik RP	GCGCGCTAATAGGGTCTTT	

Supplemental references

1. Shimazaki N, Tsai AG, & Lieber MR (2009) H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations. Mol. Cell 34:535-544.
2. Lu C, Ward A, Bettridge J, Liu Y, & Desiderio S (2015) An autoregulatory mechanism imposes allosteric control on the V(D)J recombinase by histone H3 methylation. Cell Rep. 10:29-38.