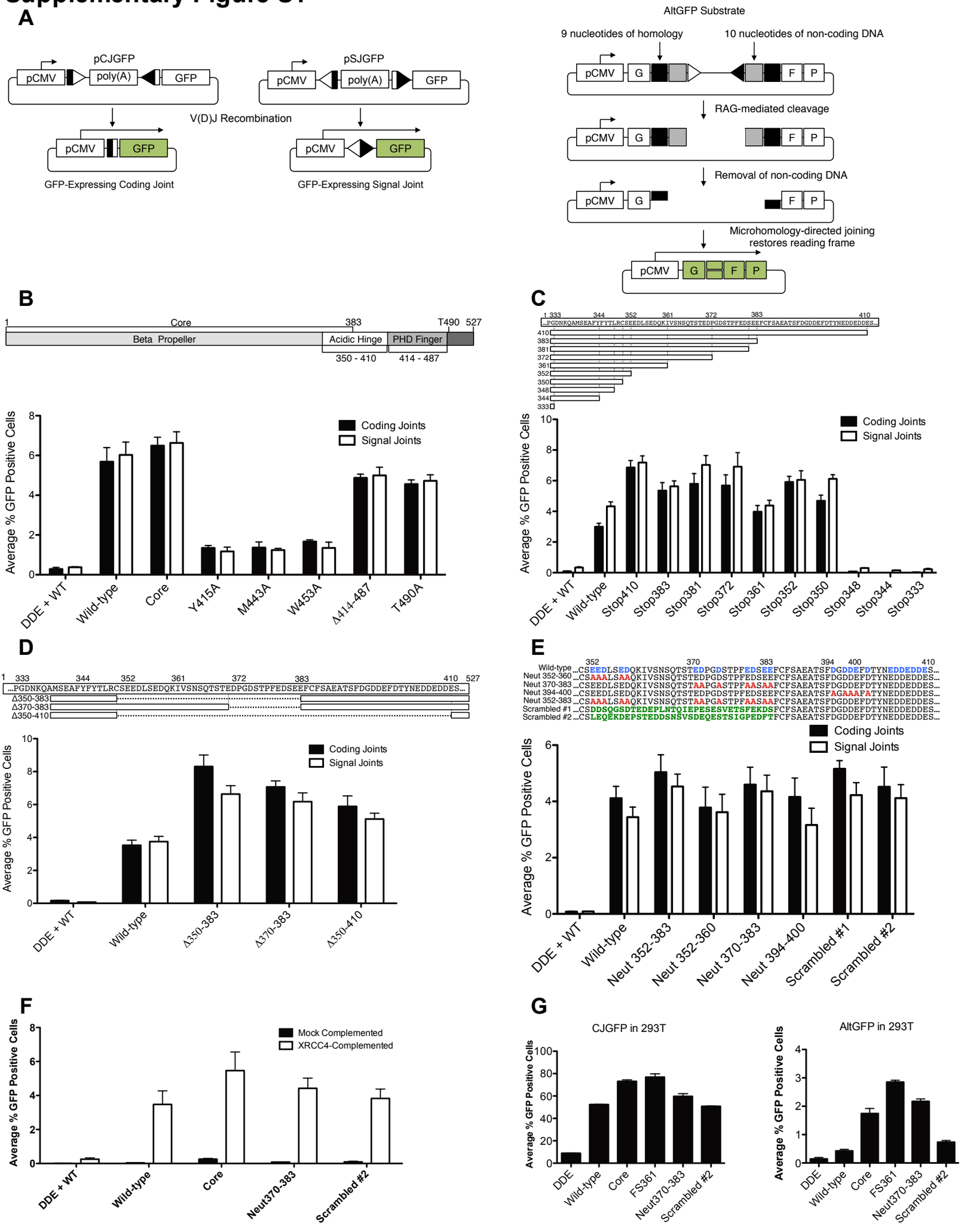
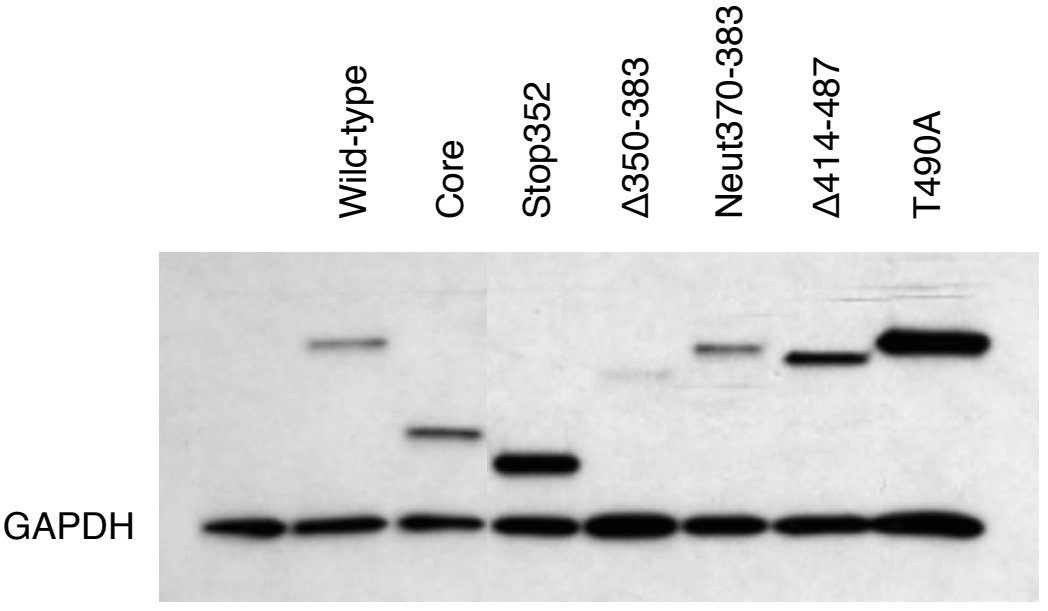
Supplementary Figure S1



Supplementary Figure S2

A



B

```
Homo sapiens            351 --AEDDTNEEQT TFTNSQTSTE DPGDSTPFED SEEFCFSAEA NSFD--GDDEFD TYNEDDEEDE SE
Pan troglodytes         351 --AEDDTNEEQT TFTNSQTSTE DPGDSTPFED SEEFCFSAEA NSFD--GDDEFD TYNEDDEEDE SE
Macaca mulatta          351 --AEEDTNEEQT TFTNSQTSTE DPGDSTPFED SEEFCFSAEA NSFD--GDDEFD TYNEDDEEDE SE
Mus musculus            351 --SEEDLSEDQK IVSNSQTSTE DPGDSTPFED SEEFCFSAEA TSFD--GDDEFD TYNEDDEDDE SV
Rattus norvegicus       351 --SEDNSSEEQK IVSNSQTSTE DAGDSTPFED SEEFCFSAEA ISFD--GDDEFD TYNEDDEDDE SV
Oryctolagus cuniculus   351 --TEDDVHEDQR TFTNSQTSTE DPGDSTPFED SEEFCFSAEA NSFD--GDDEFD TYNEDDEDDE SE
Bos taurus              351 --AEDDVNEDQI TLTSSQTSTE DPGDSTPFED SEEFCFSAEA NSFD--GDDEFD TYNEDDEEDE SE
Canis familiaris        351 --AEDDANEDQK TLANSQTSTE DPGDSTPFED SEEFCFSAEA NSFD--GDDAFD TYNEDDEEDE SE
Gallus gallus           358 NKAEEDEEEELT AQTCSQASTE DQGDSTPFED SEEFSFSAEA SSFD--VDD-ID TYNEDDEEDE SE
Xenopus boumbaensis     351 G------DNDPA LQTCSQGSTE EQEDSMPLED SEEFTFNRDG NIFD--EDT--- ---------- --
Carcharhinus plumbeus   352 G-----QLEEES QVSNSQESFT DPEECSTLED SEEFTFEVEA EGSD--EAENWD NGDEE----- --
Danio rerio             351 Q-KEQDG--EAT AQGGSQEST- DFEDSAPLED SEELYFGREP HELEYSSDVEGD TYNEEDEEDE SQ
Oncorhynchus mykiss     352 Q-KEGEGKGEDG NQVCSQEST- DFEDSAPLED SEELYFGREP HELEDSSEGEGD TYNEEDEEDE SQ
```

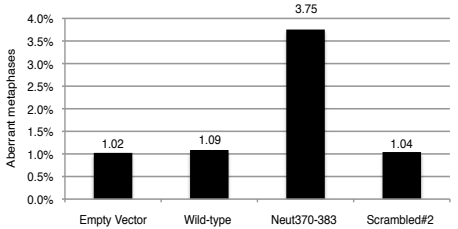
Supplementary Figure 3

A

Experiment 1	Normal	Aberrant	Cells analyzed
Empty Vector	98.98%	1.02%	391
Wild-type	98.91%	1.09%	368
Neut370-383	96.25%	3.75%	373
Scrambled#2	98.96%	1.04%	384

Statistical Analysis (Fisher's two-tailed Exact Test)

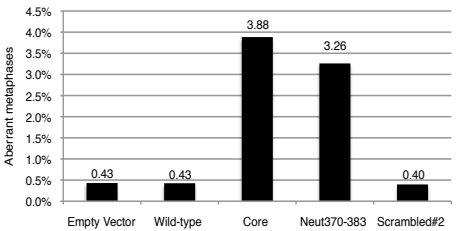
Sample A	Sample B	Probability	Significance
Wild-type	Empty Vector	1.00E+00	No
Wild-type	Neut370-383	1.20E-02	Yes
Wild-type	Scrambled#2	1.00E+00	No



Experiment 2	Normal	Aberrant	Cells analyzed
Empty Vector	99.57%	0.43%	232
Wild-type	99.57%	0.43%	235
Core	96.11%	3.88%	206
Neut370-383	96.73%	3.26%	184
Scrambled#2	99.60%	0.40%	253

Statistical Analysis (Fisher's two-tailed Exact Test)

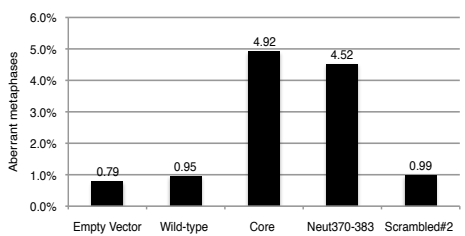
Sample A	Sample B	Probability	Significance
Wild-type	Empty Vector	1.00E+00	No
Wild-type	Core	1.40E-02	Yes
Wild-type	Neut370-383	1.20E-02	Yes
Wild-type	Scrambled#2	1.00E+00	No



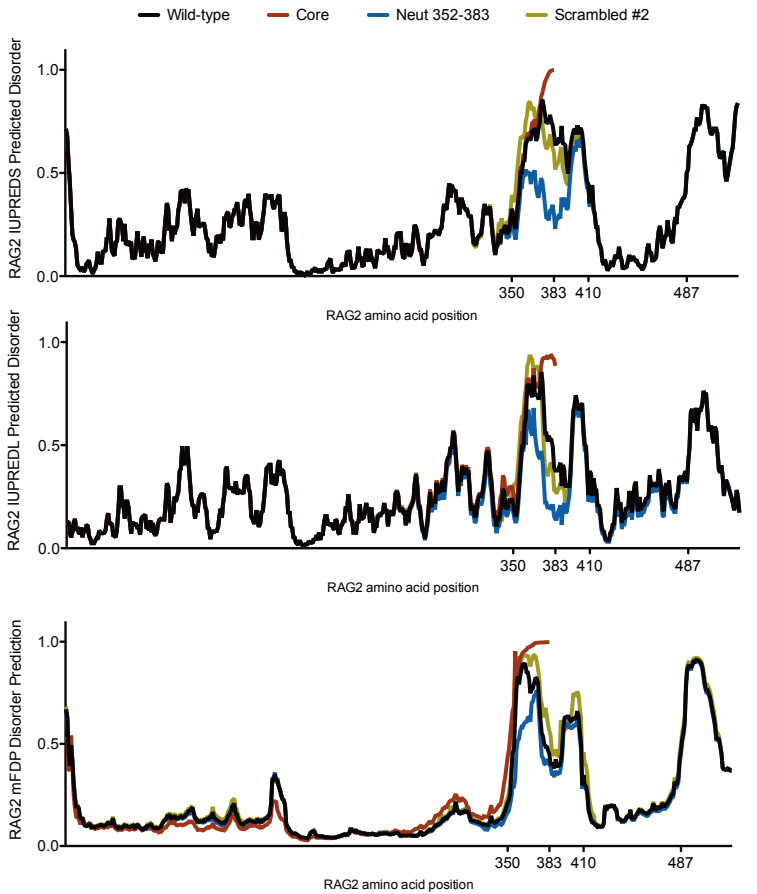
Experiment 3	Normal	Aberrant	Cells analyzed
Empty Vector	99.21%	0.79%	254
Wild-type	99.05%	0.95%	315
Core	95.08%	4.92%	325
Neut370-383	95.48%	4.52%	221
Scrambled#2	99.01%	0.99%	304

Statistical Analysis (Fisher's two-tailed Exact Test)

Sample A	Sample B	Probability	Significance
Wild-type	Empty Vector	1.00E+00	No
Wild-type	Core	3.97E-03	Yes
Wild-type	Neut370-383	1.00E-02	Yes
Wild-type	Scrambled#2	1.00E+00	No



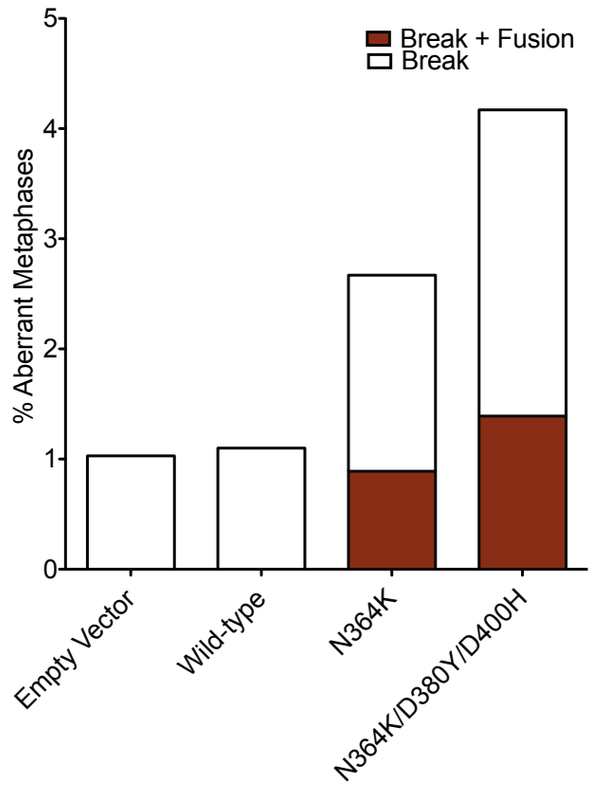
B



Supplementary Figure S4

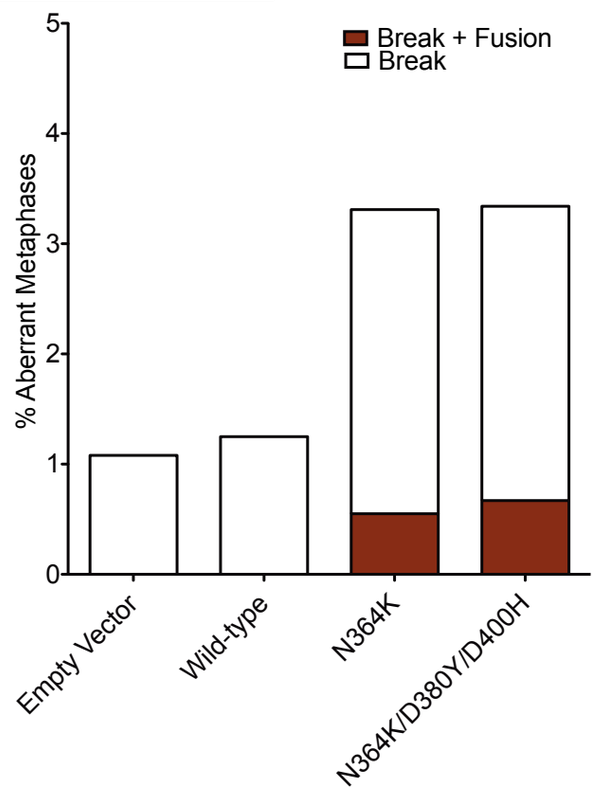
Experiment 1	Normal	Instability	Break	Break + Fusion	Cells Analyzed
Empty Vector	98.97%	1.03%	1.03%	0.00%	194
Wild-type	98.90%	1.10%	1.10%	0.00%	273
N364K	97.33%	2.67%	1.78%	0.89%	225
NDD	95.83%	4.17%	2.78%	1.39%	216

Statistical Analysis (Fisher's two-tailed Exact Test)			
Sample A	Sample B	Probability	Significance
Wild-type	Empty Vector	1.00E+00	No
Wild-type	N364K	3.10E-01	No
Wild-type	N364K/D380Y/D400H	3.86E-02	Yes



Experiment 2	Normal	Instability	Break	Break + Fusion	Cells Analyzed
Empty Vector	98.92%	1.08%	1.08%	0.00%	186
Wild-type	98.75%	1.25%	1.25%	0.00%	240
N364K	96.69%	3.31%	2.76%	0.55%	181
NDD	96.67%	3.33%	2.67%	0.67%	150

Statistical Analysis (Fisher's two-tailed Exact Test)			
Sample A	Sample B	Probability	Significance
Wild-type	Empty Vector	1.00E+00	No
Wild-type	N364K	1.80E-01	No
Wild-type	N364K/D380Y/D400H	2.69E-01	No



SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure S1. Measuring the effects RAG2 mutations on the formation of coding and signal joints, related to Figure 1.
(A) Extrachromosomal substrates used in this study for coding joints (left), signal joints (middle), and coding joints formed through alternative NHEJ (right). Quantification of FACS data from three separate experiments in CHO-K1 cells 48 hours after transfection with the indicated Rag expression vectors and the reporter substrates measuring (B) mutations to known regulatory regions, (C premature truncation mutations, as indicated, (D) internal deletion mutants, as indicated, and (E) neutralizing mutations within the acidic hinge. (F) Verification of XRCC4-deficient cells by measuring coding joint formation by complementation experiments with wild-type XRCC4. (G) Measuring coding joints (left) and AltGFP-coding joints (right) in the human 293T cell line.

Supplementary Figure S2. RAG2 mutant protein levels and conservation, related to Figures 1 and 2.
(A) A western blot of indicated HA-tagged RAG2 mutants expressed in fibroblasts. 15ug of soluble cell lysate was separated by gel-electrophoresis and transferred to a nitrocellulose membrane and exposed antibody to the HA tag for detection. Lanes were arranged for presentation; (B) Sequence alignment from indicated species of the acidic region in RAG2 performed by ClustalW. Residues indicated in black are completely conserved, while those indicated in grey are charge-conserved (Boxshade 3.21), supplementing previously published sequenced alignment data.

Supplementary Figure S3 Data from separate metaphase-FISH experiments and flexibility predictions, related to Figure 3.
(A) Three metaphase-FISH experiments showing the frequency at which the 3' and 5' *Ig*κ signals were lost in pre-B cells expressing wild-type, core, Neut370-383, and Scrambled#2 RAG2. (B) Predictions based on the IUPREDS, IUPREDL, and mFDP algorithms for Wild-type RAG2 (black), core RAG2 (red), Neut370-383 RAG2 (blue) and Scrambled#2 RAG2 (yellow).

Supplementary Figure S4 Data from separate metaphase-FISH experiments using human SVs, related to Figure 4.
Separate metaphase-FISH experiments showing the frequency at which the 3' and 5' *Ig*κ signals were lost in pre-B cells expressing wild-type, N364K, and N364K/D380Y/D400H RAG2.

EXTENDED EXPERIMENTAL PROCEDURES

Cell culture, mutagenesis and recombination assays

Chinese hamster ovary (CHO-K1 and RMP41) fibroblasts, the CHO mutant cell lines XR-1 (Xrcc4-deficient), mouse embryonic fibroblasts, and 293T cells were grown in DMEM (Invitrogen) with fetal bovine serum (10%), non-essential amino acids and penicillin-streptomycin. Pre-B-cell lines were grown in RPMI-1640 (Invitrogen) with fetal bovine serum (12%), non-essential amino acids, and penicillin-streptomycin. RAG2 point mutations and internal deletions were carried out using QuikChange II Site-Directed Mutagenesis Kit (Stratagene) or overlapping PCR, respectively, and confirmed by sequence analysis. The extra-chromosomal recombination assays were carried out as previously described (Arnal et al., 2010; Corneo et al., 2007)

Protein purification and RAG PCC stability

GST-tagged core RAG1 and the indicated untagged RAG2 proteins were prepared as described (Deriano et al., 2011; Huye et al., 2002). Two independent protein preparations were assayed for each mutant. End-release assay to measure the stability of the signal-end complexes was performed as previously described (Arnal et al., 2010; Deriano et al., 2011). Student's t-test assuming equal variance was used to calculate statistical significance.

Measuring RAG2 protein by western blot

HA-tagged RAG2 expression vectors were transfected into 293T cells. Cells were lysed, and loaded onto a 4-12% Tris-Glycine gel (Lonza) and transferred to a nitrocellulose membrane. RAG proteins were detected using a peroxidase-conjugated antibody to the HA tag (Roche). GAPDH was detected using a monoclonal anti-GAPDH-peroxidase antibody (Sigma-Aldrich).

Recombination Assay on Integrated Substrate

pMX-INV (Bredemeyer et al., 2006) was integrated into Prkdc$^{scid/scid}$ SV40-immortalized mouse embryonic fibroblasts as described previously (Deriano et al., 2009). GFP positive cells were detected by flow cytometry and analyzed with FlowJo (TreeStar, Inc.).

Hybrid Joining

Murine RAG1 and RAG2 expression constructs (encoding the indicated mutants), as well as recombination substrate pJH299 (inversional substrate) (Hesse et al., 1987) were transfected into CHO-K1 cells and hybrid joining events were scored as previously described (Deriano et al., 2009). Recombination events were directly subjected to PCR analysis with specific primers for hybrid joints (DR99-DR100) and inversional coding joints (DR99-ML68), and a region of the plasmid backbone using CMC1 (GCTGTTCGACTTACAAACACAGG) and CMC2 (GGGAAGAGGCGGTTGGG).

Generation of metaphases in RAG2-expressing pre-B cell lines

V-abl-transformed RAG2$^{-/-}$K3 pre-B cell lines were generated as previously described (Gapud et al., 2011). The indicated RAG2 mutants were cloned into the pMIT retroviral vector (pMIT-RAG2-IRES-CD90.1), and were then used to complement the V-abl-transformed RAG2$^{-/-}$ pre-B cells. Cells were sorted for the CD90.1 cell surface marker using MACS columns (Miltenyi Biotec). Quantitative real-time PCR was conducted using primers that amplify an N-terminal region of RAG2 that remains constant across all mutants tested. For separate experiments, sorted populations of complemented cells were treated with 3.0uM STI571 (Novartis) for 48 hours, washed and allowed to recover for 48 hours before they were fixed as described (Bredemeyer et al., 2006). Fixed cell suspensions were dropped onto pre-chilled glass slides

and air-dried for further analysis as described (Theunissen and Petrini, 2006). Slides were assigned random sample numbers to eliminate bias during analysis.

DNA FISH probes and DNA FISH on metaphase spreads

BAC probes for the Igκ locus were labeled by nick-translation and prepared as previously described (Hewitt et al., 2009; Skok et al., 2007). BAC RP24-507J1 and BAC RP24-218K16 were labeled with Alexa Fluor 594 and Cy3, respectively (Molecular Probes). StarFISH-concentrated mouse FITC chromosome 6 paint was prepared following supplier's instructions (Cambio). Cells were mounted in ProLong Gold (Invitrogen) containing 4′,6-diamidino-2-phenylindole (DAPI) to counterstain total DNA. Metaphase spreads were analyzed using a MetaSystems Metafer and Isis Fluorescence Imaging. A number of samples were independently re-scored by an outside investigator with fully consistent results.

Predicting disordered regions

The indicated RAG2 sequences were processed through the mFDP, IUPREDL, and IUPREDS disorder predictive algorithms which determined the level of relative disorder, as previously described (Ferron et al., 2006; Mizianty et al., 2010).

References

Arnal, S.M., Holub, A.J., Salus, S.S., and Roth, D.B. (2010). Non-consensus heptamer sequences destabilize the RAG post-cleavage complex, making ends available to alternative DNA repair pathways. Nucleic Acids Res *38*, 2944-2954.

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.

Corneo, B., Wendland, R.L., Deriano, L., Cui, X., Klein, I.A., Wong, S.Y., Arnal, S., Holub, A.J., Weller, G.R., Pancake, B.A.*, et al.* (2007). Rag mutations reveal robust alternative end joining. Nature *449*, 483-486.

Deriano, L., Chaumeil, J., Coussens, M., Multani, A., Chou, Y., Alekseyenko, A.V., Chang, S., Skok, J.A., and Roth, D.B. (2011). The RAG2 C terminus suppresses genomic instability and lymphomagenesis. Nature *471*, 119-123.

Deriano, L., Stracker, T.H., Baker, A., Petrini, J.H., and Roth, D.B. (2009). Roles for NBS1 in alternative nonhomologous end-joining of V(D)J recombination intermediates. Mol Cell *34*, 13-25.
Ferron, F., Longhi, S., Canard, B., and Karlin, D. (2006). A practical overview of protein disorder prediction methods. Proteins *65*, 1-14.
Gapud, E.J., Lee, B.S., Mahowald, G.K., Bassing, C.H., and Sleckman, B.P. (2011). Repair of chromosomal RAG-mediated DNA breaks by mutant RAG proteins lacking phosphatidylinositol 3-like kinase consensus phosphorylation sites. J Immunol *187*, 1826-1834.
Hesse, J.E., Lieber, M.R., Gellert, M., and Mizuuchi, K. (1987). Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals. Cell *49*, 775-783.
Hewitt, S.L., Yin, B., Ji, Y., Chaumeil, J., Marszalek, K., Tenthorey, J., Salvagiotto, G., Steinel, N., Ramsey, L.B., Ghysdael, J.*, et al.* (2009). RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. Nat Immunol *10*, 655-664.
Huye, L.E., Purugganan, M.M., Jiang, M.M., and Roth, D.B. (2002). Mutational analysis of all conserved basic amino acids in RAG-1 reveals catalytic, step arrest, and joining-deficient mutants in the V(D)J recombinase. Mol Cell Biol *22*, 3460-3473.
Mizianty, M.J., Stach, W., Chen, K., Kedarisetti, K.D., Disfani, F.M., and Kurgan, L. (2010). Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. Bioinformatics *26*, i489-496.
Skok, J.A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W., and Busslinger, M. (2007). Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. Nat Immunol *8*, 378-387.
Theunissen, J.W., and Petrini, J.H. (2006). Methods for studying the cellular response to DNA damage: influence of the Mre11 complex on chromosome metabolism. Methods Enzymol *409*, 251-284.

EXTENDED EXPERIMENTAL PROCEDURES

Cell culture, mutagenesis and recombination assays

Chinese hamster ovary (CHO-K1 and RMP41) fibroblasts, the CHO mutant cell lines XR-1 (Xrcc4-deficient), mouse embryonic fibroblasts, and 293T cells were grown in DMEM (Invitrogen) with fetal bovine serum (10%), non-essential amino acids and penicillin-streptomycin. Pre-B-cell lines were grown in RPMI-1640 (Invitrogen) with fetal bovine serum (12%), non-essential amino acids, and penicillin-streptomycin. RAG2 point mutations and internal deletions were carried out using QuikChange II Site-Directed Mutagenesis Kit (Stratagene) or overlapping PCR, respectively, and confirmed by sequence analysis. The extra-chromosomal recombination assays were carried out as previously described (Arnal et al., 2010; Corneo et al., 2007)

Protein purification and RAG PCC stability

GST-tagged core RAG1 and the indicated untagged RAG2 proteins were prepared as described (Deriano et al., 2011; Huye et al., 2002). Two independent protein preparations were assayed for each mutant. End-release assay to measure the stability of the signal-end complexes was performed as previously described (Arnal et al., 2010; Deriano et al., 2011). Student’s t-test assuming equal variance was used to calculate statistical significance.

Measuring RAG2 protein by western blot

HA-tagged RAG2 expression vectors were transfected into 293T cells. Cells were lysed, and loaded onto a 4-12% Tris-Glycine gel (Lonza) and transferred to a nitrocellulose membrane. RAG proteins were detected using a peroxidase-conjugated antibody to the HA tag (Roche). GAPDH was detected using a monoclonal anti-GAPDH-peroxidase antibody (Sigma-Aldrich).

Recombination Assay on Integrated Substrate

pMX-INV (Bredemeyer et al., 2006) was integrated into Prkdc$^{scid/scid}$ SV40-immortalized mouse embryonic fibroblasts as described previously (Deriano et al., 2009). GFP positive cells were detected by flow cytometry and analyzed with FlowJo (TreeStar, Inc.).

Hybrid Joining

Murine RAG1 and RAG2 expression constructs (encoding the indicated mutants), as well as recombination substrate pJH299 (inversional substrate) (Hesse et al., 1987) were transfected into CHO-K1 cells and hybrid joining events were scored as previously described (Deriano et al., 2009). Recombination events were directly subjected to PCR analysis with specific primers for hybrid joints (DR99-DR100) and inversional coding joints (DR99-ML68), and a region of the plasmid backbone using CMC1 (GCTGTTCGACTTACAAACACAGG) and CMC2 (GGGAAGAGGCGGTTGGG).

Generation of metaphases in RAG2-expressing pre-B cell lines

V-abl-transformed RAG2$^{-/-}$K3 pre-B cell lines were generated as previously described (Gapud et al., 2011). The indicated RAG2 mutants were cloned into the pMIT retroviral vector (pMIT-RAG2-IRES-CD90.1), and were then used to complement the V-abl-transformed RAG2$^{-/-}$ pre-B cells. Cells were sorted for the CD90.1 cell surface marker using MACS columns (Miltenyi Biotec). Quantitative real-time PCR was conducted using primers that amplify an N-terminal region of RAG2 that remains constant across all mutants tested. For separate experiments, sorted populations of complemented cells were treated with 3.0uM STI571 (Novartis) for 48 hours, washed and allowed to recover for 48 hours before they were fixed as described (Bredemeyer et al., 2006). Fixed cell suspensions were dropped onto pre-chilled glass slides

and air-dried for further analysis as described (Theunissen and Petrini, 2006). Slides were assigned random sample numbers to eliminate bias during analysis.

DNA FISH probes and DNA FISH on metaphase spreads

BAC probes for the Igκ locus were labeled by nick-translation and prepared as previously described (Hewitt et al., 2009; Skok et al., 2007). BAC RP24-507J1 and BAC RP24-218K16 were labeled with Alexa Fluor 594 and Cy3, respectively (Molecular Probes). StarFISH-concentrated mouse FITC chromosome 6 paint was prepared following supplier's instructions (Cambio). Cells were mounted in ProLong Gold (Invitrogen) containing 4′,6-diamidino-2-phenylindole (DAPI) to counterstain total DNA. Metaphase spreads were analyzed using a MetaSystems Metafer and Isis Fluorescence Imaging. A number of samples were independently re-scored by an outside investigator with fully consistent results.

Predicting disordered regions

The indicated RAG2 sequences were processed through the mFDP, IUPREDL, and IUPREDS disorder predictive algorithms which determined the level of relative disorder, as previously described (Ferron et al., 2006; Mizianty et al., 2010).

References

Arnal, S.M., Holub, A.J., Salus, S.S., and Roth, D.B. (2010). Non-consensus heptamer sequences destabilize the RAG post-cleavage complex, making ends available to alternative DNA repair pathways. Nucleic Acids Res *38*, 2944-2954.

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.

Corneo, B., Wendland, R.L., Deriano, L., Cui, X., Klein, I.A., Wong, S.Y., Arnal, S., Holub, A.J., Weller, G.R., Pancake, B.A.*, et al.* (2007). Rag mutations reveal robust alternative end joining. Nature *449*, 483-486.

Deriano, L., Chaumeil, J., Coussens, M., Multani, A., Chou, Y., Alekseyenko, A.V., Chang, S., Skok, J.A., and Roth, D.B. (2011). The RAG2 C terminus suppresses genomic instability and lymphomagenesis. Nature *471*, 119-123.

Deriano, L., Stracker, T.H., Baker, A., Petrini, J.H., and Roth, D.B. (2009). Roles for NBS1 in alternative nonhomologous end-joining of V(D)J recombination intermediates. Mol Cell *34*, 13-25.

Ferron, F., Longhi, S., Canard, B., and Karlin, D. (2006). A practical overview of protein disorder prediction methods. Proteins *65*, 1-14.

Gapud, E.J., Lee, B.S., Mahowald, G.K., Bassing, C.H., and Sleckman, B.P. (2011). Repair of chromosomal RAG-mediated DNA breaks by mutant RAG proteins lacking phosphatidylinositol 3-like kinase consensus phosphorylation sites. J Immunol *187*, 1826-1834.

Hesse, J.E., Lieber, M.R., Gellert, M., and Mizuuchi, K. (1987). Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals. Cell *49*, 775-783.

Hewitt, S.L., Yin, B., Ji, Y., Chaumeil, J., Marszalek, K., Tenthorey, J., Salvagiotto, G., Steinel, N., Ramsey, L.B., Ghysdael, J.*, et al.* (2009). RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. Nat Immunol *10*, 655-664.

Huye, L.E., Purugganan, M.M., Jiang, M.M., and Roth, D.B. (2002). Mutational analysis of all conserved basic amino acids in RAG-1 reveals catalytic, step arrest, and joining-deficient mutants in the V(D)J recombinase. Mol Cell Biol *22*, 3460-3473.

Mizianty, M.J., Stach, W., Chen, K., Kedarisetti, K.D., Disfani, F.M., and Kurgan, L. (2010). Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. Bioinformatics *26*, i489-496.

Skok, J.A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W., and Busslinger, M. (2007). Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. Nat Immunol *8*, 378-387.

Theunissen, J.W., and Petrini, J.H. (2006). Methods for studying the cellular response to DNA damage: influence of the Mre11 complex on chromosome metabolism. Methods Enzymol *409*, 251-284.