Supplemental Information

Supplemental Figure Legends

Figure S1. Kinetics of parasitemia and AID^{GFP} expression, related to Figure 1.

A- Parasitemia over time in wild type mice inoculated with *Pc*. Left: representative flow cytometry plots showing parasitized erythrocytes (gate). Right: summary plot of parasitemia over time. Mean value with standard deviation is shown; at least five mice for each time point.

B- Representative flow cytometry plots of *Plasmodium* infected AID^{GFP} spleens at different time points after inoculation. See also Figure 1C. At least two mice for each time point.

Figure S2. *In vitro* validation of the ROSA^{AIDer} and ROSA^{erISCEI} alleles, related to Figure 2.

A- Class switch recombination *in vitro*. ROSA$^{AIDer/+}$ B cells and controls were stimulated for 4 days with LPS and IL4, in the presence or absence of 4OH-Tamoxifen, prior to flow cytometry for IgG1. The gate indicates IgG1 switched B cells, FSC is forward scatter for cell size.

B- Western blot detection of erISCEI. B cells of the indicated genotypes were cultured for 4 days in the presence of LPS, IL4, and anti-CD180. The erISCEI protein was detected with anti-HA antibodies recognizing this N-terminal tag in erISCEI (arrow). B cells infected by an erISCEI expressing retrovirus were used as positive control.

C- erISCEI induces the recombination between I-SceI target sequences on the same chromosome. The ROSA^{erISCEI} transgene was bred to AID$^{-/-}$ and to the IgH$^{I\text{-}96k}$ allele to generate ROSA$^{erISCEI/+}$IgH$^{I\text{-}96k/+}$AID$^{-/-}$. The IgH$^{I\text{-}96k/+}$ allele bears two I-SceI sites 96kb apart at the IgH locus on chromosome 12. Recombination between the I-SceI sites is achieved following I-SceI mediated cleavage, and can be detected by PCR using primers adjacent to the sites (see diagram and (Bothmer et al., 2010)). Shown is an ethidium bromide stained agarose gel with PCR products on cells cultured for four days under the indicated conditions.

D- erISCEI induces the recombination between I-SceI target sequences on different chromosomes. Similar to C, The ROSA^{erISCEI} transgene was bred to AID$^{-/-}$ and to the Myc^{I} and IgH^{I} alleles to generate ROSA$^{erISCEI/+}$Myc$^{I/I}$IgH$^{I/I}$AID$^{-/-}$. The Myc^{I} and IgH^{I} alleles each bear a single I-SceI site at the *c-myc* (chromosome 15) and *Igh* (chromosome 12) loci. Recombination between I-SceI sites can be achieved following I-SceI mediated cleavage, and can be detected by PCR using primers adjacent to the sites (see diagram and (Robbiani et al., 2008)). Shown is an ethidium bromide stained agarose gel with PCR products on cells cultured for four days in the presence of 1μM 4OH-Tamoxifen. Recombination between I-SceI sites was confirmed by direct sequencing of the amplicons.

E- Relative mRNA expression for the indicated genes, normalized with Tubulin (Hogenbirk et al., 2013). Numbers indicate the fold difference between samples (*in vivo*/*in vitro*).

Figure S3. Features of translocations within hotspots of viral integration at ERFS in malaria GC B cells, related to Figure 3.

A- Proportion of genic rearrangements within ERFS.

B- Frequency of rearrangements within ERFS at genes with various levels of transcription. Empty bars represent AID deficient sample. Dashed line represents the expected frequency based on random model.

Figure S4. Genomic damage by AID in malaria germinal centers, related to Figure 4.

A- Mutational analysis by MutPE-seq. Coverage (top) and relative mutation frequency (bottom) at each position for the indicated amplicons. Blue is $AID^{-/-}$, red is IgκAID transgene. See also Figure 4C.

B- Proportion of genic rearrangements within off-target AID hotspots.

C- Frequency of rearrangements within off-target AID hotspots at genes with various levels of transcription. Dashed line represents the expected frequency based on random model.

Figure S5. Lymphoid development in AID proficient and deficient $CD19^{cre/+}p53^{lox/lox}$ mice and impaired ability to control *Plasmodium* in $AID^{-/-}$ mice, related to Figure 5.

A- Flow cytometry of lymphoid tissues shows no substantial developmental differences in $CD19^{cre/+}p53^{lox/lox}$ compared to controls. Two experiments.

B- qPCR on genomic DNA shows efficient deletion of the floxed p53 locus in purified $CD19^{cre/+}p53^{lox/lox}$ B cells ($CD19^{+}$ fraction). Wild type control is shown alongside. Mean with standard deviation of two independent experiments.

C- Flow cytometry of lymphoid tissues shows no substantial developmental differences in $CD19^{cre/+}p53^{lox/lox}AID^{-/-}$ compared to controls. Two experiments.

D- Survival of immunodeficient NRG mice inoculated with 5-10 mio $CD19^{cre/+}p53^{lox/lox}AID^{-/-}$ splenocytes from sick mice. At signs of distress mice were sacrificed and analyzed. Splenomegaly was present in all, and lymphoma was confirmed by flow cytometry.

E- Hematologic values in $CD19^{cre/+}p53^{lox/lox}AID^{-/-}$ mice (12-14 months old) are consistent with increased erythrolysis as a consequence of chronic *Plasmodium* infection. Controls are same age uninfected mice of the same genotype.

F- Parasitemia over time upon a single dose of *Pc* administration. Mice with $p53^{lox/lox}$ are also $CD19^{cre/+}$. Representative of three independent experiments.

G- Hematologic values in representative mice 17 weeks after *Pc* infection.

H- Splenomegaly in AID deficient mice 21 weeks upon infection with *Pc*. Extramedullary hematopoiesis was confirmed histologically (data not shown).

I- Decreased survival in the absence of AID in p53 proficient mice infected with *Pc*.

Figure S6. Lymphomas in $CD19^{cre/+}p53^{lox/lox}$ mice, related to Figure 6 and Table 1.

A- Survival of *Plasmodium* infected, SRBC immunized, and control mice. All mice are also $CD19^{cre/+}$.

B- Representative flow cytometry plots of $CD19^{cre/+}p53^{lox/lox}$ lymphomas. L4 and L23 are post-GC B cell lymphomas that arose in *Plasmodium* infected mice. L22 is a pre-GC B cell lymphoma that developed in uninfected control. Numbers are the percentage of cells within the respective gate. FSC is forward scatter for cell size.

C- Distribution of lymphoma phenotypes in SRBC immunized $CD19^{cre/+}p53^{lox/lox}$ mice.

Figure S7. Deep-sequencing of lymphoma L23, related to Figure 6.

A- Deep-sequencing of malaria-associated lymphoma L23. Left: schematic of paired-end library preparation. Middle: ethidium bromide stained agarose gel with sonicated and adapter-ligated tumor DNA (smear). Right: Agilent Bioanalyzer results of excised and pre-amplified ~537-575bp and ~576-612bp bands to verify size and purity before paired-end sequencing. The ~537-575bp bands were used for deep-sequencing on Illumina platform (see Methods for details).

B- Ethidium bromide stained agarose gels with PCR products to confirm and identify translocation breakpoints in L23. Left panel: amplicon generated with primers 5-GTGGAGGTGTATGGGGTGTAGAC-3 and 5-CCTCAGTCACCGTCTCCTCAGGTA-3 to amplify *c-myc*/*Igh* of T(12;15). Middle panel (from left to right): amplicon for T(18;3) with primers 5-GCTAGGCCCAAGAATAGCCTC-3 and 5-CAGCTGCGACTGAACCATTG-3; amplicon for T(11;15) with primers 5-CCTAGCCAGTGGACCTTGTC-3 and 5-GAGGATGAGGGGACAGACAG-3; amplicons for T(8;17) with primers 5-GCACCGTGCGTTCCGCGTCTC-3 and 5-GGCTCTGGAGGTATTTAGGG-3 or with primers 5-GCCTTTCTCCCCCAACCCCCC-3 and 5-CAGGAGCTATGAGATCCCGG-

3. Right panel: DNA quality control PCR at the 53bp1 locus with primers 5-GTGCCTCATTGTTGGGGAGAG-3 and 5- TGTGTGGTTCACCTTCTCTCTATGG-3. DNAs from tumor L4, wild type tail (wt), and water (w) are controls.

C- Breakpoint analysis of translocations in lymphoma L23. Fifty nucleotides surrounding each breakpoint (middle) are shown, with the homology to germline sequence (top and bottom) indicated by vertical bars. Microhomologies are yellow; insertions green.

Supplemental Information

Extended Experimental Procedures

Mutational analysis by Paired-End deep-sequencing (MutPE-seq)

To determine the frequency of mutations by AID (Figure 4C), 50 ng of genomic DNA from FACS-sorted malaria GC B cells (3 weeks post infection) were amplified by PCR and prepared for deep-sequencing. The first round of PCR amplification was performed with Phusion polymerase (annealing temperature of 55C, 30 cycles) and locus-specific primers for Switch μ (5'-TCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TCT CTG AGT GCT TCT AAA ATG CG and 5'-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TTC ACC CCA ACA CAG CGT AGC), *c-myc* (5'-TCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT AAA GGG GAA GGG AAA AAC C and 5'-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TCC AAA CGC AAA AGG TAA TCC [c-myc (A)]; 5'-TCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ACA GGG ATG TGA CCG ATT CGT TG and 5'-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TAC CTC CCT TCT ACA CTC TAA ACC [c-myc (B)]), and near *Ly6e* (5'-TCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT AAG AAG AGC CTT ATT GGA GTG G and 5'-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TAA CTC AGA AAT CCA CCT ACC TC). Five-ten ng of DNA from the first reaction was used as template for the second reaction (annealing temperature of 65 C, 10 cycles) to introduce sequencing adapters and genotype-specific barcodes (underlined). Primers were: 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC and 5'-CAA GCA

GAA GAC GGC ATA CGA GAT ATTGGC GTG ACT GGA GTT CAG ACG TGT G ($AID^{-/-}$) or 5'-CAA GCA GAA GAC GGC ATA CGA GAT TACAAG GTG ACT GGA GTT CAG ACG TGT G (IgκAID). Gel-extracted amplicons were pooled and sequenced from both ends with MiSeq 600. Mutations present in only one of the paired reads were considered a sequencing artifact and discarded.

TC-seq and computational analysis (continuation)

The observed difference in unique rearrangements between AID deficient and AID proficient TC-seq datasets *in vivo* is statistically significant ($p < 0.0001$ with the binomial test, Figure 2D). The initial, overall DNA damage is expected to be comparable between AID deficient and proficient samples, with a slight increase in the AID proficient sample due to AID mutator activity. A combination of factors may however be responsible for the observed difference in the number of uniquely captured events in favor of the $AID^{-/-}$ sample: (1) More efficient capture in $AID^{-/-}$, due to higher levels of erISCEI. Cells from the AID deficient sample have two erISCEI alleles ($ROSA^{erISCEI/erISCEI}$), while the AID proficient sample has only one, the other ROSA allele being taken by the AIDer transgene ($ROSA^{erISCEI/AIDer}$). Two alleles produce (in theory) double the amount of enzyme, which is expected to cleave more efficiently at the I-SceI target site. More cutting will offer more opportunities for the cleaved I-SceI site to recombine with a concomitant DNA break elsewhere in the genome, leading to a unique rearrangement that we then capture by TC-seq. Twice the erISCEI amounts may not precisely double the cleavage or recombination efficiencies, as the relationship may not be linear, but a significantly higher number of rearrangements is expected in the AID deficient sample

for this reason. (2) Less efficient library preparation in $AID^{+/+}$, due to a smaller number of rearrangements to start with (directly linked to point 1 above). In the generation of TC-seq libraries, a series of enzymatic reactions are alternated with DNA purifications steps (gel-, column-, or beads-purification). At each round of purification DNA is lost, and the less rearrangement-associated DNA you start with, the more (in proportion) you loose. This issue results in a more pronounced loss of diversity/richness in the libraries built on fewer rearrangements, here affecting more the AID proficient sample. (3) *In vivo* selection is more likely to eliminate cells suffering AID damage. We have shown that AID causes DNA damage (DNA breaks and mutations) not randomly, but with a preference for genes that are important for B cell identity/development. As a consequence, GC B cells suffering AID damage at those key genes are more likely to be selected against *in vivo*, leaving behind those with fewer rearrangements. Deleterious AID mutations at IgV genes would also be selected against. This phenomenon would be irrelevant for cells cultured *in vitro*.

Pathology

Peripheral blood analysis, tissue processing, and histopathologic evaluation were performed at the Laboratory of Comparative Pathology, Memorial Sloan-Kettering Cancer Center.

qPCR

For p53 deletion, $CD19^{+}$ splenic B cells were enriched using anti-CD19 microbeads (Miltenyi Biotech) and the $CD19^{-}$ fractions served as control. Genomic DNA from 2

million cells was amplified using primers for p53 (5-GGTTAAACCCAGCTTGACCA-3 and 5-GGAGGCAGAGACAGTTGGAG-3) and normalized with *Gapdh (*5-CGAAGGTGGAAGAGTGGGAG-3 and 5-TGAAGCAGGCATCTGAGGG-3).

Table S2. Histology of B cell lymphomas in *P. chabaudi* infected CD19$^{cre/+}$p53$^{lox/lox}$ mice, related to Table 1.

Tumor ID	Histologic evaluation (spleen)
L4	The majority of cells consist of large neoplastic lymphocytes, with scant basophilic cytoplasm and large polygonal to reniform vesicular nuclei. Nucleoli are indistinct. Anisocytosis and anisokaryosis are approximately 3-fold, and frequent mitoses are present.
L9	Normal tissue architecture is effaced and replaced by sheets of a relatively homogeneous population of large round to polygonal lymphoid cells having moderately well-delineated cell margins, scant amphophilic cytoplasm, and large polygonal to reniform vesicular nuclei with variably prominent nucleoli. Anisocytosis and anisokaryosis are approximately 3-fold, and mitoses average 3 per 40x field. Occasional tingible-body macrophages are scattered throughout; however a distinct "starry sky" appearance is not present.
L23	The majority of cells consist of large neoplastic lymphocytes, with scant basophilic cytoplasm and large polygonal to reniform vesicular nuclei. Nucleoli are indistinct. Anisocytosis and anisokaryosis are approximately 3-fold, and frequent mitoses are present. A few aggregates of coarse granular brown to black pigment are visible, consistent with hemozoin.
L24	Normal tissue architecture is regionally effaced and replaced by sheets of cells that are similar in morphology to those described in lymphoma L9; however anisocytosis and anisokaryosis are approximately 5-fold, and mitoses average 4 per 40x field. Tingible-body macrophages are not apparent.
L33	ND
L52	White pulp areas are multifocally moderately to markedly expanded and coalescing with a relatively homogeneous population of large cells that are similar in appearance and mitotic index to those described in tumor L24. Occasional tingible-body macrophages are scattered throughout; however a distinct "starry sky" appearance is not present.
L62	White pulp areas are multifocally moderately expanded and coalesce in many areas with a relatively homogenous population of large cells that are similar in morphology and mitotic index to those described in lymphoma L9. Tingible-body macrophages are not apparent.
L68	White pulp areas are multifocally moderately expanded and coalesce in many areas with a relatively homogenous population of small round to polygonal lymphoid cells having relatively well-delineated cell margins, scant amphophilic cytoplasm, and large polygonal to reniform hyperchromatic nuclei with indistinct nucleoli. Anisocytosis and anisokaryosis are approximately 2-fold, and mitoses average 1 per 10 40x fields. Tingible-body macrophages are not apparent.
L31	Normal tissue architecture is effaced and replaced by sheets of a relatively homogeneous population of large cells that are similar in morphology and mitotic index to those described in sample L24. Tingible-body macrophages are not apparent.
L39	White pulp areas are multifocally moderately expanded and coalesce in many areas with a relatively homogenous population of large cells that are similar in morphology and mitotic index to those described in tumor L9. Tingible-body macrophages are not apparent.
L69	White pulp areas are multifocally moderately expanded and coalesce in some areas with a relatively homogenous population of large cells that are similar in morphology and mitotic index to those described in lymphoma L9. Tingible-body macrophages are not apparent.

Table S3. Karyotype and phenotype of lymphomas in control uninfected $CD19^{cre/+}p53^{lox/lox}$ mice, related to Table 1. Translocations are clonal if present in all metaphases. In italic are sporadic translocations and bolded are clonal, reciprocal translocations. At least 20 metaphases analyzed for each tumor. Macroscopic tumor dissemination at necropsy: S = spleen; MLN = mesenterial lymph nodes; sbLN = submandibular lymph nodes; oLN = axillary, thoracal or inguinal (other) lymph nodes; L = liver or other metastasis. ND is not determined.

Tumor ID	Translocation	Tumor Phenotype	S	MLN	sbLN	oLN	L
L8	ND	$CD3^+$	+	+	+	+	-
L10	*T(17;18) T(16;2)*	Mixed $CD3^+$ and $CD19^+$	+	-	-	+	-
L36	ND	$CD3^+$	+	+	-	-	-
L7	no metaphases	$CD19^+$ IgM^- IgD^- IgK^+	+	+	+	-	-
L64	ND	$CD19^-$ IgM^- IgD^- IgK^- $CD138^+$ $CD38^+$ FAS^+ $GL7^-$	+	+	-	+	-
L1	*T(12;15) T(12;3)*	$CD19^+$ IgM^- IgD^- IgK^- $CD138^-$ $CD38^+$ FAS^+ $GL7^-$	+	+	+	+	-
L19	no metaphases	$CD19^+$ IgM^+ IgD^- IgK^+ $CD138^-$ $CD38^+$ FAS^+ $GL7^{dim}$	+	-	-	-	-
L22	T(12;6) **T(4;14)**	$CD19^+$ IgM^+ IgD^- IgK^+ $CD138^-$ $CD38^+$ FAS^+ $GL7^-$	+	-	+	+	-
L25	*T(5;7) T(15;7) T(11;14)*	$CD19^+$ IgM^+ IgD^- IgK^+ $CD138^-$ $CD38^+$ FAS^+ $GL7^{dim}$	+	-	-	+	-
L53	T(16;10) T(2;5)	$CD19^+$ IgM^+ IgD^- IgK^+ $CD138^-$ $CD38^+$ $FAS^{+/-}$ $GL7^-$	+	-	-	-	+
L57	T(5;13) T(14;1) T(4;5)	$CD19^+$ IgM^+ IgD^- IgK^+ $CD138^-$ $CD38^+$ FAS^+ $GL7^{dim}$	+	-	-	+	-
L58	ND	$CD19^+$ IgM^+ IgD^- IgK^+ $CD138^-$ $CD38^+$ FAS^+ $GL7^+$	+	+	-	-	-
L63	ND	$CD19^+$ IgM^+ IgD^+ IgK^+ $CD138^-$ $CD38^+$ FAS^+ $GL7^{dim}$	+	-	+	-	-
L85	ND	$CD19^-$ $B220^+$ IgM^- IgD^- IgK^- $CD138^-$ $CD38^+$ FAS^+ $GL7^+$	+	+	-	+	-
L86	ND	$CD19^+$ IgM^- IgD^- IgK^- $CD138^-$ $CD38^+$ $FAS^{+/-}$ $GL7^{dim}$	+	+	+	+	-

Table S4. Phenotype of lymphomas in $CD19^{cre/+}p53^{lox/lox}$ mice immunized with SRBC, related to Table 1. Macroscopic tumor dissemination at necropsy: S = spleen; MLN = mesenterial lymph nodes; sbLN = submandibular lymph nodes; oLN = axillary, thoracal or inguinal (other) lymph nodes; L = liver or other metastasis. ND is not determined.

Tumor ID	Tumor Phenotype	S	MLN	sbLN	oLN	L
L55	$CD3^+$	+	+	+	+	-
L65	Mixed $CD3^+$ and $CD19^+$	-	-	-	+	-
L27	$CD19^+$ IgM^- IgD^- IgK^+ $CD138^+$ $CD38^+$ FAS^{dim} $GL7^+$	+	+	-	-	-
L54	$CD19^+$ IgM^- IgD^- IgK^+ $CD138^+$ $CD38^+$ FAS^+ $GL7^{dim}$	+	-	-	+	-
L2	$CD19^+$ IgM^- IgD^- IgK^- $CD138^-$ $CD38^+$ FAS^- $GL7^-$	-	+	+	+	-
L42	$CD19^+$ IgM^- IgD^- IgK^- $CD138^-$ $CD38^+$ FAS^{dim} $GL7^+$	+	-	+	-	-
L87	$CD19^+$ IgM^+ IgD^- IgK^+ $CD138^-$ $CD38^+$ FAS^+ $GL7^-$	+	+	+	+	+
L48	$CD19^-$ $B220^-$ $CD3^-$	+	-	-	-	-
L56	$CD19^-$ $B220^-$	+	-	+	-	-
L59	$CD19^-$ $B220^-$ $CD3^-$	+	-	+	+	-
L88	ND	+	-	-	-	-

Supplemental References

Bothmer, A., Robbiani, D.F., Feldhahn, N., Gazumyan, A., Nussenzweig, A., and Nussenzweig, M.C. (2010). 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. J Exp Med *207*, 855-865.

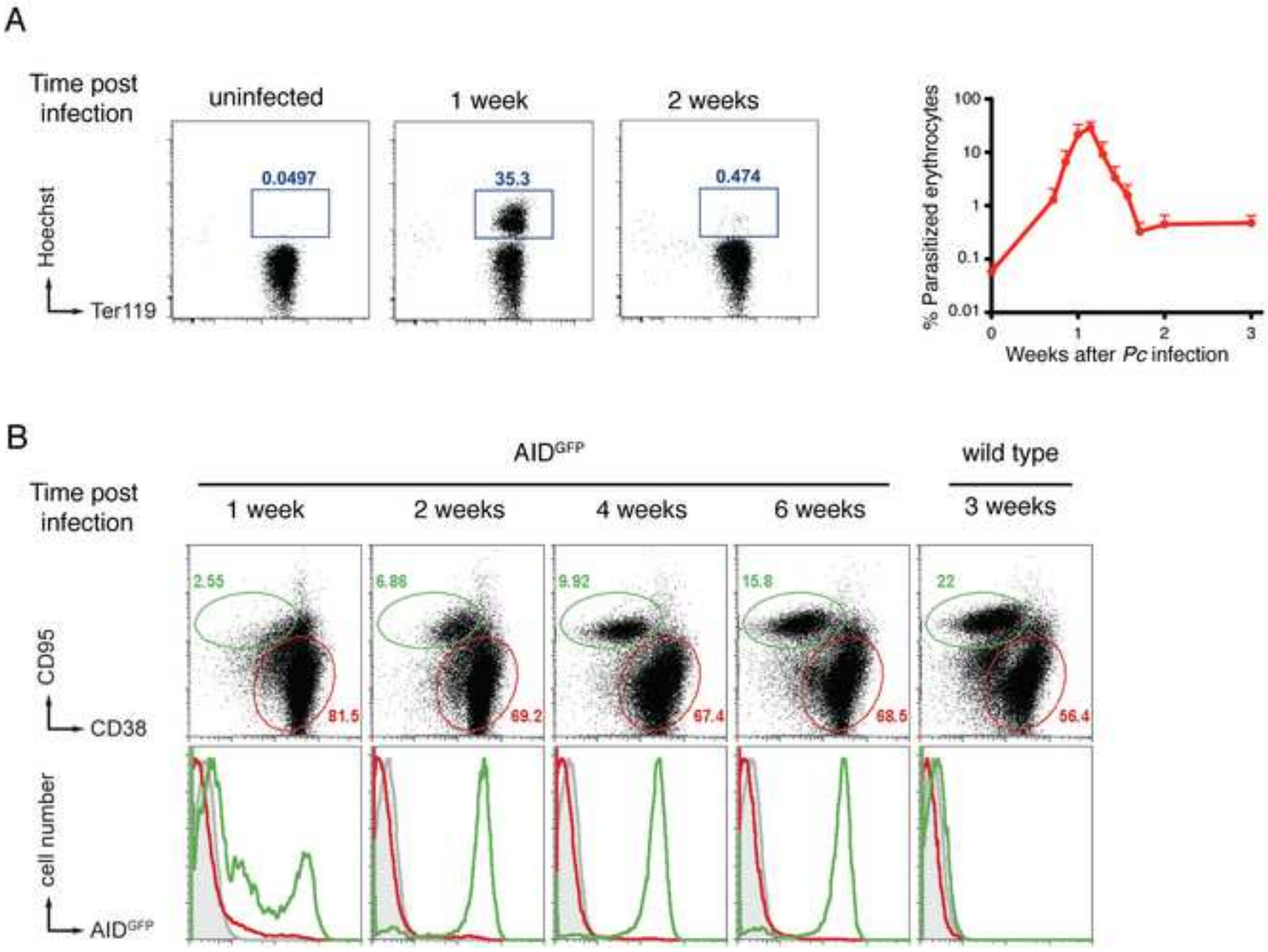
Hogenbirk, M.A., Heideman, M.R., Velds, A., van den Berk, P.C., Kerkhoven, R.M., van Steensel, B., and Jacobs, H. (2013). Differential programming of B cells in AID deficient mice. PloS one *8*, e69815.

Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A.*, et al.* (2008). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. Cell *135*, 1028-1038.

Supplemental Figure 1

Click here to download Supplemental Figure: Figure S1.tif

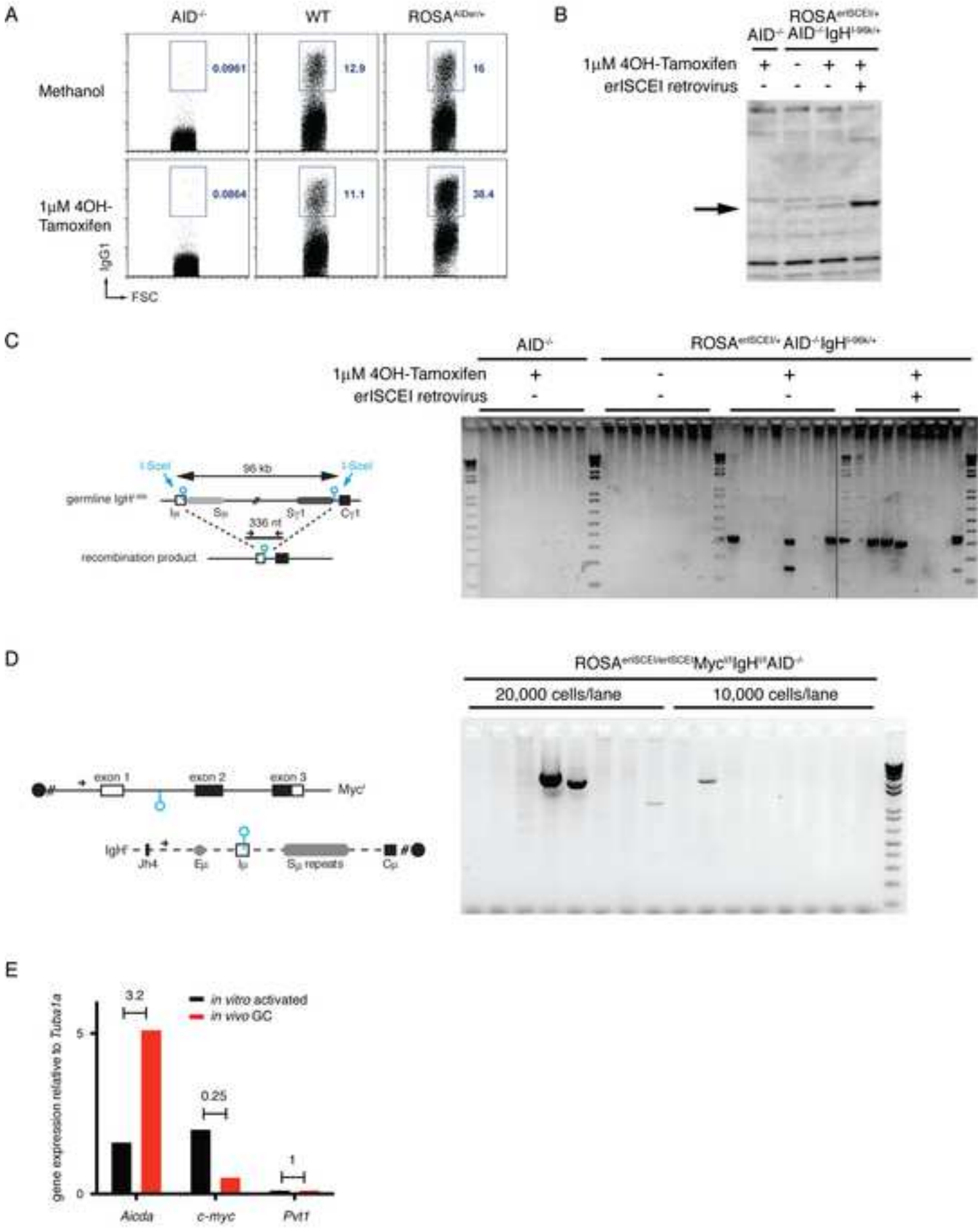
Figure S1



Supplemental Figure 2

Click here to download Supplemental Figure: Figure S2.tif

Figure S2



Supplemental Figure 3
Click here to download Supplemental Figure: Figure S3.tif

Figure S3

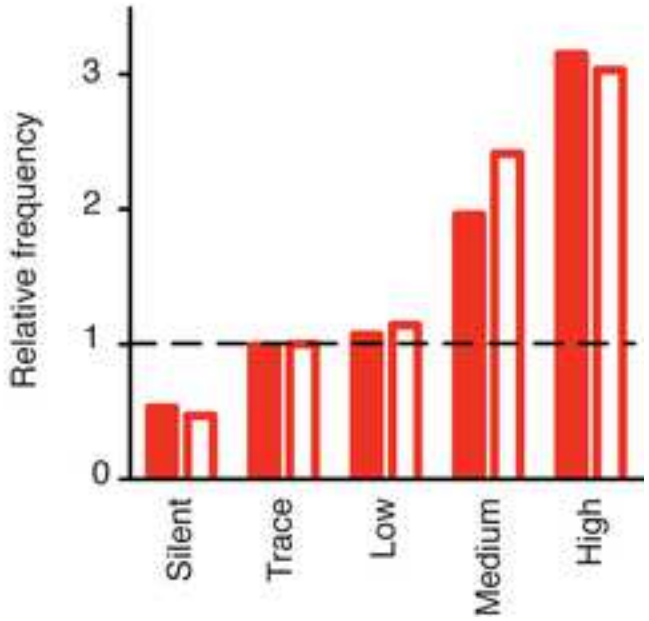
A

Captured rearrangements at ERFS hotspots

		total	genic	% genic
in vivo malaria GC	AID proficient	736	499	67.8
	AID deficient	2514	1907	75.9

p <0.0001 over predicted random distribution of 40.1%
(binomial test)

B

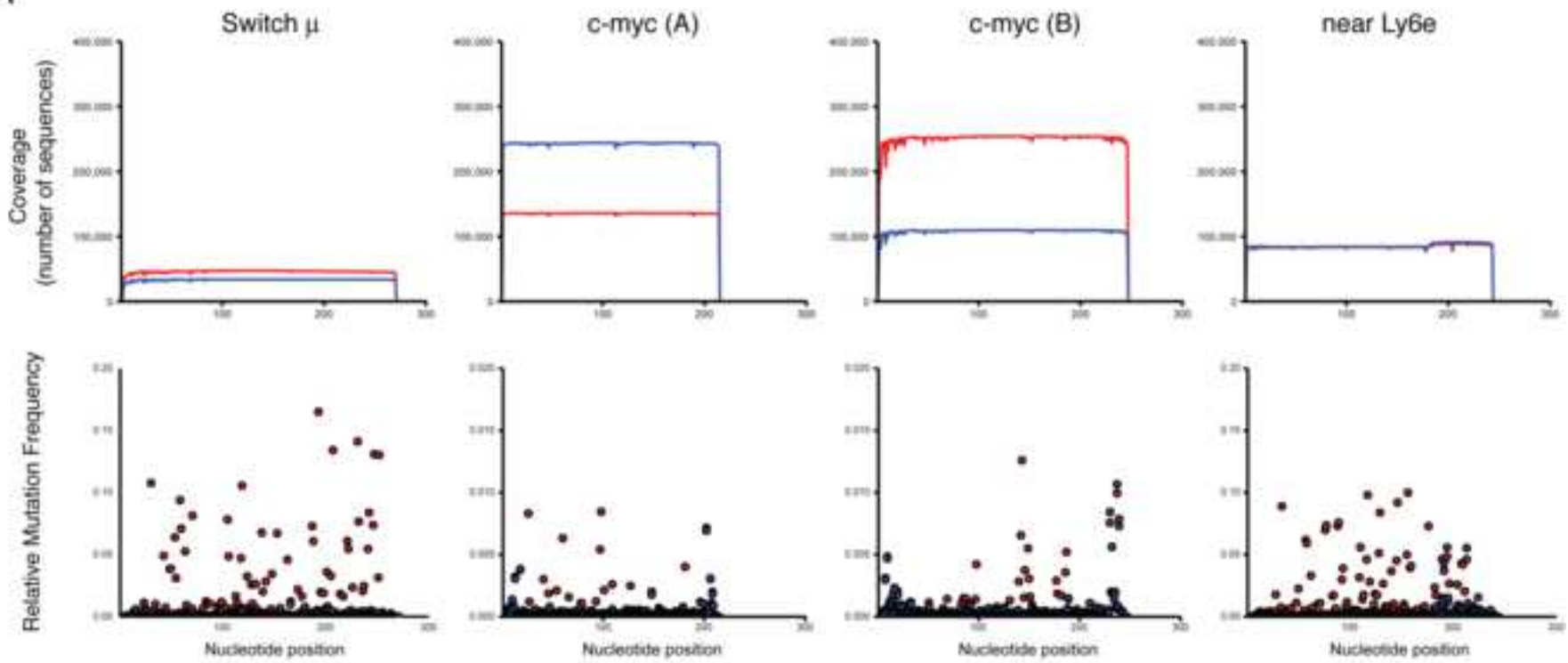


Supplemental Figure 4

Click here to download Supplemental Figure: Figure S4.tif

Figure S4

A



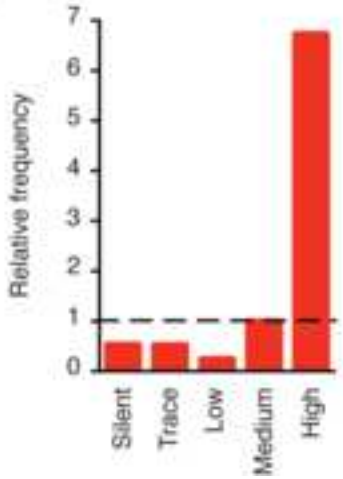
B

Captured rearrangements at AID hotspots

	total	genic	% genic
in vivo malaria GC	160	81	50.6

p <0.01 over predicted random distribution of 40.1% (binomial test)

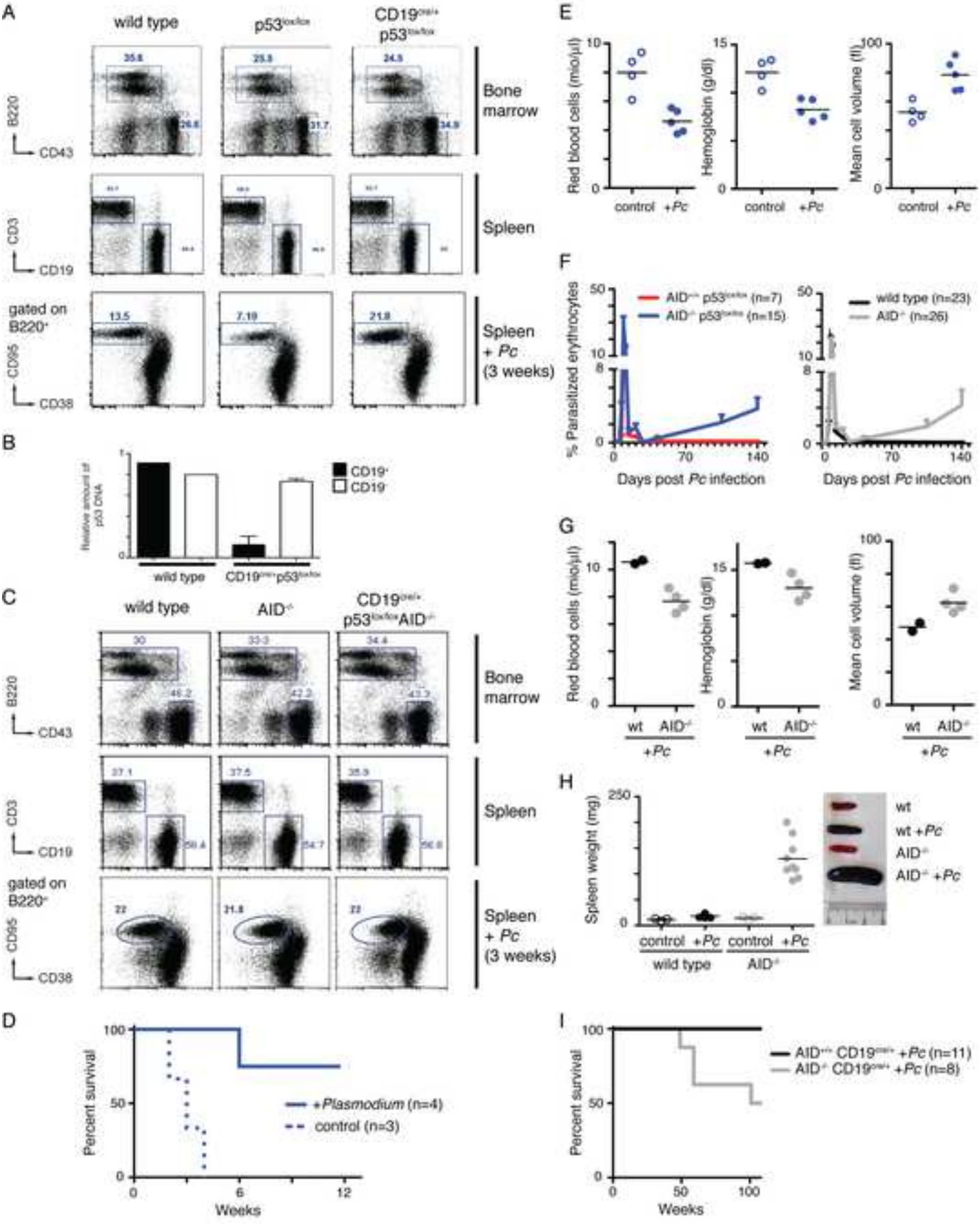
C



Supplemental Figure 5

Click here to download Supplemental Figure: Figure S5.tif

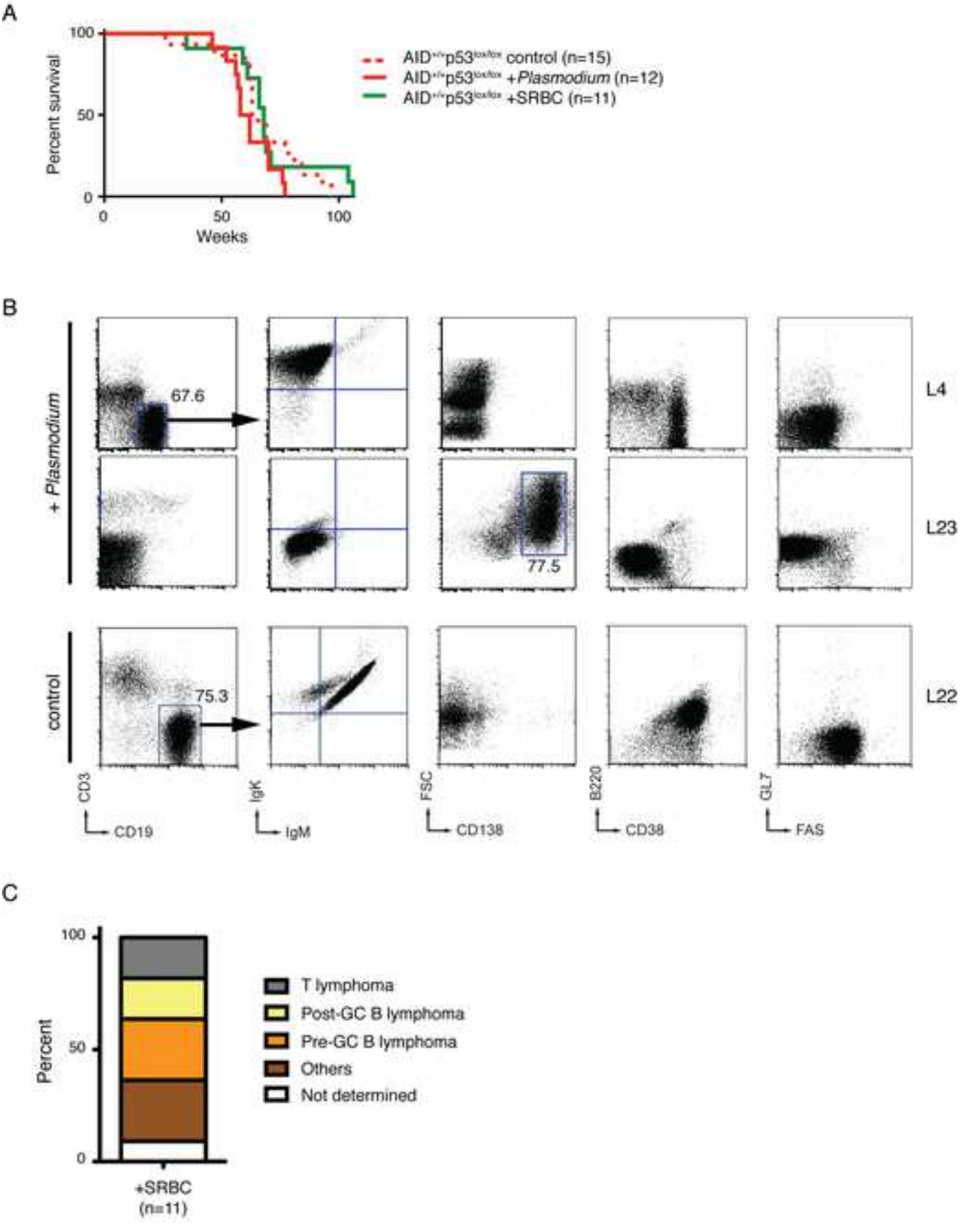
Figure S5



Supplemental Figure 6

Click here to download Supplemental Figure: Figure S6.tif

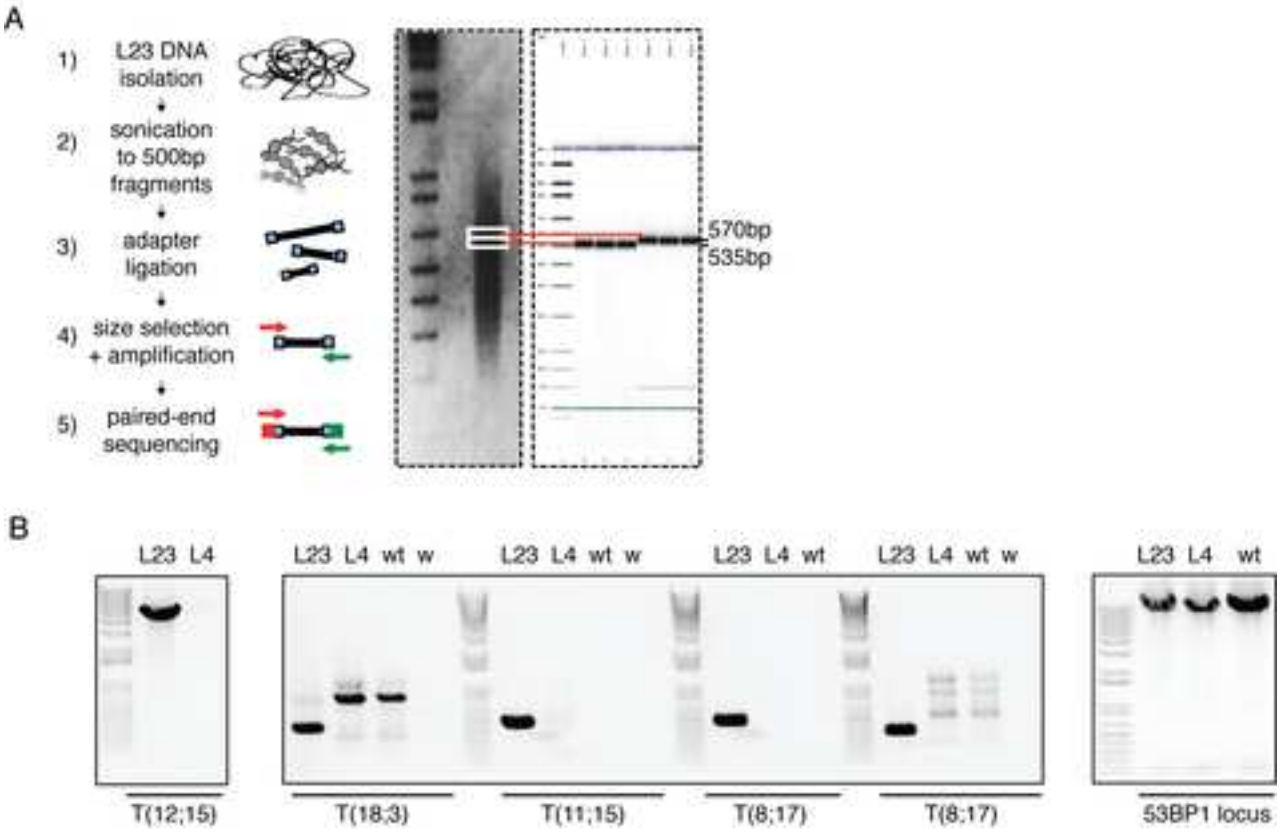
Figure S6



Supplemental Figure 7

Click here to download Supplemental Figure: Figure S7.tif

Figure S7



Supplemental Table S1

Click here to download Supplemental Movies and Spreadsheets: Table S1.xlsx

Supplemental Table S5

Click here to download Supplemental Movies and Spreadsheets: Table S5.xlsx