SUPPLEMENTAL INFORMATION

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Generation and characterization of c-myc$^{1xI\text{-}SceI}$ and of ROSA$^{I\text{-}SceI\text{-}GR}$ mice (refers to Figure 1).

(A) Schematic map of the targeting strategy for insertion of a single I-SceI site into *c-myc* intron 1. The c-myc$^{1xI\text{-}SceI}$ allele was generated by homologous recombination into 129/Sv (TC1) ES cells of the indicated construct. Position of probe used for ES cell screening by Southern blotting is indicated. (B) Southern blot analysis of DNA from a targeted clone before and after Neo-deletion. DNA was digested with EcoRI and probed with the 3' probe indicated in (A). (C) Efficiency of cutting of the c-myc$^{1xI\text{-}SceI}$ allele was tested by metaphase FISH. Splenic primary B cells were activated for CSR by αCD40/IL4 and infected with either control (C) or I-SceI (I) retrovirus. Metaphase spreads were prepared at day 4 and analyzed with probes flanking the *c-myc* locus (see Suppl. Experimental Procedures). Numbers of c-myc breaks are indicated for 3 independent c-myc$^{1xI\text{-}SceI}$ mice and two c-myc$^{25xI\text{-}SceI}$ mice as positive control. (D) The ROSA$^{I\text{-}SceI\text{-}GR}$ allele was generated by homologous recombination into 129/Sv (TC1) ES cells of the indicated construct. The I-SceI-GR-IRES-tdT cassette was targeted into the intron 1 of the Rosa26 locus. Position of the probe used for ES cell screening by Southern blotting is indicated. (E) Southern blot analysis of DNA from a targeted clone before and after Neo-deletion. DNA was digested with EcoRI and probed with the 5' probe indicated in (D). (F) Primary B cells from ROSA$^{I\text{-}SceI\text{-}GR}$ targeted mice express the Tomato red fluorescent protein as detected by cytofluorometry. B cells were activated for 4 days with αCD40/IL4. CSR was evaluated by staining with αB220 and αIgG1 antibodies. (G) Efficient I-SceI target sequence cutting by the I-SceI-GR protein. B cells from c-myc$^{25xI\text{-}SceI}$/ROSA$^{I\text{-}SceI\text{-}GR}$ mice were activated for 4 days by αCD40/IL4. Addition of TA to the culture at day 1 induced cutting and rejoining in the 25x I-SceI cassette, as shown by disappearance of the full-length PCR product and presence of a smear of shorter fragments in day 4 DNA samples from treated cells. Control PCR amplification was

conducted on plasmids containing either 25 or 1 I-SceI site(s). A diagram showing position of PCR primers is presented on top.

Figure S2. Detailed genome-wide maps of translocations from HTGTS libraries (refers to Figure 2).

In each map single translocation junctions are represented by dots located at the corresponding chromosomal position. Dot scale, color key and ideogram orientation are as given for Figure 2. Data to the right or left of chromosome plot are identified at the top of each genome map. Panels A-F show libraries made from I-SceI sites in the *c-myc* locus (chr15). Panels G and H show libraries with the $\Delta S\gamma 1^{2xI\text{-}SceI}$ cassette at the *IgH* locus (chr12). The $\Delta S\gamma 1^{2xI\text{-}SceI}$ data are from a pool of HTGTS libraries from 3 (WT) or 2 ($AID^{-/-}$) independent mice (See Table S2). (A) Comparison of the two methods used for HTGTS library generation; the pool of 6 WT c-$myc^{25xI\text{-}SceI}$ circularization-PCR libraries (13782 junctions) is shown on the left; the pool of 4 WT c-$myc^{25xI\text{-}SceI}$ adapter-PCR libraries (42470 junctions) is shown on the right. (B) Comparison of adapter-PCR libraries from two independent c-$myc^{25xI\text{-}SceI}$ mice. (C) Comparison of HTGTS libraries from c-$myc^{25xI\text{-}SceI}$ and c-$myc^{1xI\text{-}SceI}$ B cells. (D) Comparison of HTGTS libraries from I-SceI-infected c-$myc^{25xI\text{-}SceI}$ and TA-induced c-$myc^{25xI\text{-}SceI}$/$ROSA^{I\text{-}SceI\text{-}GR}$ WT B cells. (E) Comparison of representative c-$myc^{25xI\text{-}SceI}$/$AID^{-/-}$ and c-$myc^{25xI\text{-}SceI}$/WT HTGTS libraries. (F) Comparison of HTGTS libraries from two additional c-$myc^{25xI\text{-}SceI}$/$AID^{-/-}$ mice. (G) $\Delta S\gamma 1^{2xI\text{-}SceI}$ translocations sorted by orientation, as in Figure 2. (H) Comparison of $\Delta S\gamma 1^{2xI\text{-}SceI}$ translocation maps from WT (right) and $AID^{-/-}$ (left) libraries. Arrows indicate the position of hotspots, as defined in Figure 4.

Figure S3. Origin and distribution of (+) and (-) orientation junctions in HTGTS libraries (refers to Figure 2).

(A-H) Diagrams showing outcome of different types of joining events identified in HTGTS libraries following I-SceI-mediated DSBs in *c-myc* (A-D) and *IgH* (E-H) loci. A, E: single DSB and resection; B, F: intrachromosomal translocations to a second DSB 3' to the sequencing primers; C, G: intrachromosomal translocations to a second DSB 5' to the sequencing primers;

D, H: interchromosomal translocations. I-SceI substrates are shown as yellow boxes. Sequencing primers are indicated. (I and J) Graphs showing the distribution of junctions in the chromosomal (+) (blue bars) and (-) (red bars) orientation in c-myc$^{25xI\text{-}SceI}$ (I) and $\Delta S\gamma1^{2xI\text{-}SceI}$ (J) HTGTS libraries. Data are presented as average of three HTGTS libraries ± SEM in I and of two HTGTS libraries in J.

Figure S4. Details of chr12 and chr15 translocation clusters in HTGTS libraries (refers to Figure 3).

(A) Distribution of translocation junctions around the break-site (chr15) in the pooled c-myc$^{25xI\text{-}SceI}$/AID$^{-/-}$HTGTS library. Top: 250 kb around the breaksite, indicated by the red line; Bottom: 2.5 Mb around the breaksite. (+) and (-) orientation junctions are plotted on the top and bottom of the chromosome diagrams, respectively. (B and C) Distribution of translocation junctions around the breaksite (chr12) in $\Delta S\gamma1^{2xI\text{-}SceI}$/WT (A) and AID$^{-/-}$ (B) HTGTS library pools. Top: 250 kb around the break-site, represented by the yellow box (indicating Sγ1 deletion) and red asterisk (indicating I-SceI sites); Bottom: 2.5 Mb around the breaksite. (+) and (-) orientation junctions are plotted on the bottom and top of the chromosome diagrams, respectively. (D) Details of junctions mapping to the 20 kb region around the break-site in the pooled $\Delta S\gamma1^{2xI\text{-}SceI}$/WT (top) and AID$^{-/-}$ (bottom) HTGTS libraries. Given that Sγ1 is deleted in the $\Delta S\gamma1^{2xI\text{-}SceI}$ allele, junctions in the Sγ1 region represent trans-chromosomal events. In all panels, color codes for translocation clusters are shown in the legend. Dot size, position of centromere (red oval) and telomere (green rectangle) and orientation of the sequencing primer are indicated.

Figure S5. TSS analysis of c-myc$^{25xI\text{-}SceI}$/AID$^{-/-}$ and c-myc$^{25xI\text{-}SceI}$/WT HTGTS libraries including *IgH* switch region hotspots (refers to Figure 4).

Translocation junctions were grouped into 100bp bins and plotted according to the nearest TSS as defined in Figure 4B and 4C. GRO-seq from B cells was used to classify active and inactive promoters. Asterisks indicate translocation junctions from genes containing I-SceI target sites. (A) Overlay of active TSS-mapped translocation junctions from both WT and AID$^{-/-}$ B cells excluding 1 Mb around the *c-myc* breaksite and *IgH* S regions. (B) Same composite as

(A), but with all AID-dependent and I-SceI hotspots removed. (C and D) Junctions within 1 Mb around the *c-myc* break-site were excluded. Traces representing active promoters are in red and those containin inactive promoters are in blue. Upper panel: 40kb region around the TSS. Lower panel: 6kb region around the TSS Our current analyses have not revealed a clear enrichment for the known AID WRC target sequence in and around non-*IgH* AID translocation hotspots near TSSs. Additional analyses are ongoing to search for other potential AID-dependent translocation hotspot motifs.

Figure S6. Translocations originating from the c-myc$^{25xI\text{-}SceI}$ breaksite correlate with transcription genome-wide (refers to Figure 6).

Translocation density maps from pooled c-myc$^{25\ I\text{-}SceI}$/WT and c-myc$^{25xI\text{-}SceI}$AID$^{-/-}$ HTGTS libraries (each bar represent an independent translocation junction) are aligned with GRO-seq maps (combined sense and antisense nascent RNA signals) for chromosomes 1-10, 12-14, 16, 18, 19, and X using the UCSC genome browser. Chromosomes with their respective gene densities (genes in blue with black borders) are displayed below the GRO-seq maps. Chromosomal orientation from left to right is centromere (C) to Telomere (T).

Figure S7. Translocations originating from the ΔSγ1$^{2xI\text{-}SceI}$ breaksite correlate with transcription (refers to Figure 6).

Translocation density maps from the pooled ΔSγ1$^{2xI\text{-}SceI}$/WT HTGTS library (each bar represent an independent translocation junction) are aligned with GRO-seq maps (combined sense and antisense nascent RNA signals) for chromosomes 4, 11, 12, 15, and 17 using the UCSC genome browser. Gene densities (blue gene boxes with black borders) are indicated on the chromosome diagrams. Chromosomal orientation from left to right is centromere (C) to Telomere (T).

SUPPLEMENTAL TABLES

Table S1. Micro-homology (MH) usage at translocation junctions.

		c-myc$^{1xI\text{-}SceI}$ retro-I-SceI		c-myc$^{25xI\text{-}SceI}$ retro-I-SceI		c-myc$^{25xI\text{-}SceI}$ ROSA$^{I\text{-}SceI\text{-}GR}$	
		n	**%**	**n**	**%**	**n**	**%**
Non-breaksite[a]	**TOTAL JUNCTIONS**	503		537		195	
	INSERTIONS	94	18.7	138	25.7	42	21.5
	OTHERS[c]	409	81.3	399	74.3	153	78.5
	direct	91	22.2	109	27.3	38	24.8
	MH<5	288	70.4	244	61.2	98	64.1
	MH 5-10	28	6.8	37	9.3	17	11.1
	MH>10	2	0.5	9	2.3	0	0.0
		n	**%**	**n**	**%**	**n**	**%**
Breaksite[b]	**TOTAL JUNCTIONS**	1066		1359		547	
	INSERTIONS	304	28.5	475	35.0	123	22.5
	OTHERS[c]	762	71.5	884	65.0	424	77.5
	direct	80	10.5	108	12.2	44	10.4
	MH<5	629	82.5	685	77.5	326	76.9
	MH 5-10	53	7.0	91	10.3	54	12.7
	MH>10	NA		NA		NA	

[a]non-breaksite: junctions at genomic locations other than at the breaksite

[b]breaksite: junctions at +/- 1Mb from the c-myc I-SceI breaksite

[c]others: junctions not containing insertions; insertions were excluded from calculation of percentages of junctions with no MH (direct) or with MHs of different length.

Table S2. Mice used for generation of HTGTS libraries.

c-myc$^{1x\ I\text{-}SceI}$/WT

Mouse #	Mouse ID	sex	age	% CSR	% infection	Total n of translocations			Hits on chr15		Hits on all chrs (chr15 excluded)	
						ad	cir	tot	n	%	n	%
#1	183	F	8 mo	43	74	6042	3046	9088	7034	77.40	2054	22.60
#2	2	M	3 mo	24	87	nd	549	549	443	80.69	106	19.31
#3	9	M	2 mo	35	63.5	nd	4303	4303	3423	79.55	880	20.45
TOTAL						6042	7898	13940	10900	78.19	3040	21.81

c-myc$^{25x\ I\text{-}SceI}$/ROSA$^{I\text{-}SceI\text{-}GR}$

Mouse #	Mouse ID	sex	age	% CSR	% infection	Total n of translocations			Hits on chr15		Hits on all chrs (chr15 excluded)	
						ad	cir	tot	n	%	n	%
#1	3	M	1.5 mo	12.6	na	nd	1089	1089	868	79.71	221	20.29
#2	8	M	1.5 mo	10.8	na	nd	1100	1100	891	81.00	209	19.00
#3	16	M	2 mo	9.2	na	nd	735	735	557	75.78	178	24.22
#4	17	M	3.5	46	na	nd	393	393	302	76.84	91	23.16
#5	433,436	F	4.5 mo	9	na	9137	nd	9137	7810	85.48	1327	14.52
TOTAL						9137	3317	12454	10428	83.73	2026	16.27

c-myc$^{25x\ I\text{-}SceI}$/WT

Mouse #	Mouse ID	sex	age	% CSR	% infection	Total n of translocations			Hits on chr15		Hits on all chrs (chr15 excluded)	
						ad	cir	tot	n	%	n	%
#1	488	M	3 mo	32	60	12479	1778	14257	11497	80.64	2760	19.36
#2	484,485	M,F	3 mo	31	61	14426	nd	14426	12007	83.23	2419	16.77
#3	311	F	6.5 mo	17	63	14365	1065	15430	13539	87.74	1891	12.26
#4	248,249	F	2 mo	33	70	nd	3942	3942	2991	75.88	951	24.12
#5	253	M	5 mo	41	45	1200	1778	2978	2354	79.05	624	20.95
#6	749	M	4 mo	22	86	nd	3708	3708	3014	81.28	694	18.72
#7	858	F	1.5 mo	37	67	nd	1511	1511	1258	83.26	253	16.74
TOTAL						42470	13782	56252	46660	82.95	9592	17.05

c-myc$^{25x\ I\text{-}SceI}$/AID$^{-/-}$

Mouse #	Mouse ID	sex	age	% CSR	% infection	Total n of translocations			Hits on chr15		Hits on all chrs (chr15 excluded)	
						ad	cir	tot	n	%	n	%
#1	409	M	4 mo	1.25	74	13980	3177	17157	15000	87.43	2157	12.57
#2	487	F	3 mo	1.04	68	13149	2233	15382	13322	86.61	2060	13.39
#3	494	F	3 mo	1.15	62	9911	nd	9911	8582	86.59	1329	13.41
#4	342	M	3 mo	0.5	37	1245	2225	3470	2911	83.89	559	16.11
TOTAL						38285	7635	45920	39815	86.71	6105	13.29

ΔSγ1$^{25x\ I\text{-}SceI}$/WT

Mouse #	Mouse ID	sex	age	% CSR	% infection	Total n of translocations			Hits on chr15		Hits on all chrs (chr15 excluded)	
						ad	cir	tot	n	%	n	%
#1	186	F	1.5 mo	20	48	nd	1675	1675	1568	93.61	107	6.39
#2	188	F	8 mo	13	74	5190	972	6162	5673	92.06	489	7.94
#3	958	F	2.5 mo	27	84	nd	1542	1542	1425	92.41	117	7.59
TOTAL						5190	4189	9379	8666	92.40	713	7.60

ΔSγ1$^{25x\ I\text{-}SceI}$/AID$^{-/-}$

Mouse #	Mouse ID	sex	age	% CSR	% infection	Total n of translocations			Hits on chr15		Hits on all chrs (chr 15 excluded)	
						ad	cir	tot	n	%	n	%
#1	165	M	2 mo	0.1	40	nd	2742	2742	2561	93.40	181	6.60
#2	278	F	7 mo	0.3	77	3081	1350	4431	4115	92.87	316	7.13
TOTAL						3081	4092	7173	6676	93.07	497	6.93

Table S3. Number of hits within hotspots and in c-myc$^{25xI\text{-}SceI}$ WT and AID-/- HTGTS libraries

Gene/ intergenic region	n of hits in hotspots[a]		Binomial test (p-value)	n of total hits[b]		Chr	Genomic sequence (mm9, UCSC browser)	Size (bp)	
	WT	AID$^{-/-}$		WT	AID$^{-/-}$				
Iμ-Sμ-Cμ	446	12	2.20E-16 (p<0.001)	446	12	chr12	114,656,789-114,665,418	8,629	
Iε-Sε-Cε	327	0	2.20E-16 (p<0.001)	327	0	chr12	114,506,320-114,515,995	9,675	
Iγ1-Sγ1-Cγ1	259	8	2.20E-16 (p<0.001)	259	8	chr12	114,563,470-114,577,590	14,120	
Il4ra	35	1	6.33E-09 (p<0.001)	35	1	chr7	132,695,796-132,722,988	27,193	
Gm12493	26	1	1.32E-05 (p<0.001)	26	1	chr4	44,976,178-44,983,040	6,862	
Pim1	25	1	2.39E-05 (p<0.001)	25	1	chr17	29,627,990-29,632,404	4,415	
Aff3	19	1	5.65E-05 (p<0.001)	19	1	chr1	38,234,172-38,721,800	487,629	
J Kappa	14	0	0.000591 (p<0.001)	14	0	chr6	70,672,513-70,676,431	3,918	
Fcer2a	13	0	0.001074 (p<0.01)	13	0	chr8	3,681,737-3,694,174	12,438	
Cd83	30	6	0.001945 (p<0.01)	30	6	chr13	43,880,476-43,898,501	18,026	
Arid5a	11	0	0.003649 (p<0.01)	11	0	chr1	36,364,578-36,380,874*	16,297	
Fli1	10	1	0.006825 (p<0.01)	11	1	chr9	32,229,793-32,348,953	119,161	
chr4:44725759	12	0	0.01081 (p<0.05)	12	0	chr4	44,725,759-44,727,010	1,251	
Clec2d	8	0	0.02447 (p<0.05)	8	0	chr6	129,130,633-129,136,553	5,921	
Il21r	6	0	0.04339 (p<0.05)	8	0	chr7	132,746,943-132,777,084	30,142	**statistically significant** (p<0.05)
Bcl2l1	6	0	0.04339 (p<0.05)	6	0	chr2	152,606,404-152,657,418	51,015	
Dync1h1	6	0	0.04339 (p<0.05)	6	0	chr12	111,839,605-111,905,154	65,550	
Lrrc33	6	0	0.04339 (p<0.05)	6	0	chr16	32,142,911-32,165,562	22,652	
Mef2c	6	0	0.04339 (p<0.05)	6	0	chr13	83,643,033-83,806,684	163,652	
Socs2	6	0	0.04339 (p<0.05)	6	0	chr10	94,874,676-94,879,455*	5368	
Tnfaip3	6	0	0.04339 (p<0.05)	6	0	chr10	18,720,717-18,735,216*	14,500	
Itpk1	1	6	0.0467 (p<0.05)	1	6	chr12	103,806,793-103,943,079	136,287	
Bcl11a	9	2	0.134	9	2	chr11	23,978,056-24,073,558*	95,503	
Rapgef1	13	4	0.1448	13	4	chr2	29,475,240-29,595,883*	120,644	
Nup62-il4i1	8	2	0.2081	8	2	chr7	52,071,740-52,096,173	24,434	
Grap	6	1	0.2509	6	1	chr11	61,466,823-61,486,279	19,457	
Traf1	6	1	0.2509	6	1	chr2	34,798,778-34,817,292	18,515	
Sh3bp5	7	2	0.3192	7	2	chr14	32,187,150-32,249,219	62,070	
Zfp608	7	2	0.3192	7	2	chr18	55,047,702-55,149,567	102,136	
Pax5	9	3	0.3812	9	3	chr4	44,544,378-44,723,312	178,935	
Hivep3	6	2	0.4811	8	2	chr4	119,487,283-119,808,016	320,734	**≥3-fold enriched**

Gucy2f	8	3	0.3766	8	3	chrX	138,515,703-138,631,474	117,648	Not Significant
Rad51l1	7	3	0.5348	7	3	chr12	80,398,269-80,915,677	517,409	
Sfi1[c]	8	8	0.6149	8	8	chr11	3,031,853-3,093,466	61,614	
Ebf1	6	3	0.7425	8	5	chr11	44,431,636-44,818,674	387,039	
miR-715 region[c]	18	13	1	18	13	chr17	39,980,196-39,985,774	5,206	
Mpdu1	2	6	0.07648	2	6	chr11	69,470,206-69,476,144	5,939	I-SceI site
Mmp24	45	45	0.1355	45	45	chr2	155,601,080-155,644,102	43,023	
Scd2	4	8	0.1403	4	8	chr19	44,368,166-44,381,352	13,187	
Apbb1	19	21	0.2013	19	21	chr7	112,706,998-112,730,049*	23,052	
Fermt2	19	19	0.3288	19	19	chr14	46,078,467-46,149,740	71,274	
Kirrel3	14	15	0.3477	14	15	chr9	34,296,316-34,843,892	547,577	
chr1:31667457	10	10	0.5028	10	10	chr1	31,667,457-31,668,744	1,287	
chr1:137751395	17	15	0.5952	17	15	chr1	137,751,395-137,753,879	2,484	
Cxcr5	5	0	0.07848	5	0	chr9	44,319,870-44,334,504	14,635	5-hit hotspots & special regions
Laptm5	5	0	0.07848	5	0	chr4	130,469,249-130,492,063	22,815	
Mad1l1	5	0	0.07848	6	0	chr5	140,484,643-140,797,506	312,864	
Man1a	5	0	0.07848	5	0	chr10	53,625,839-53,795,602	169,764	
Rab35	5	0	0.07848	5	0	chr5	116,081,996-116,097,167	15,172	
Cd44	5	1	0.4115	5	1	chr2	102,651,300-102,741,822*	90,523	
Cflar	5	1	0.4115	5	1	chr1	58,770,130-58,815,726	45,597	
Dock10	5	1	0.4115	5	1	chr1	80,497,648-80,755,128	257,481	
Plekha2	5	1	0.4115	5	1	chr8	26,149,617-26,212,283	62,667	
Cdh4	5	2	0.7065	7	2	chr2	179,177,183-179,634,080	456,898	
Itpkb	5	2	0.7065	5	2	chr1	182,260,607-182,353,790	93,184	
Prkca	5	2	0.7065	6	4	chr11	107,794,701-108,205,202	410,502	
Sept9	5	3	1	5	3	chr11	117,060,975-117,223,639*	162,665	
Fnbp1	5	3	1	5	3	chr2	30,881,726-30,997,528**	115,803	
Itpr1	1	5	0.08912	2	5	chr6	108,163,090-108,501,108	338,019	
Sipa1l3	1	5	0.08912	2	5	chr7	30,105,397-30,290,479	185,083	
Zmiz1	3	5	0.2936	3	5	chr14	26,278,671-26,486,233	207,563	
9130019P16Rik	3	5	0.2936	3	5	chr6	54,219,675-54,380,215	160,541	
Plcd3	4	5	0.506	4	5	chr11	102,931,610-102,962,972	31,363	
Ig λ	7	3	0.5348	7	3	chr16	19,026,951-19,260,937	233, 986	
V kappa	2	0	NA	19	11	chr6	67,505,630-70,672,513	3.1E+06	

[a]Derived from 250kb hotspot bins
[b]Total hits present
[c]These may be false hotspots; they involve repeated sequences that have not yet been annotated in the mouse reference genome, see text
* variant 1 (the longest isoform)
** variant 3 (the longest isoform)

Table S4. Direct PCR assay for translocations between c-myc$^{25xI\text{-}SceI}$ and AID-dependent hotspots genes

	WT				AID$^{-/-}$			
	# 247		# 311		# 409		# 494	
	C	I	C	I	C	I	C	I
Il4ra	<0.17	0.83	<0.17	0.5	<0.17	<0.17	<0.17	<0.17
Il21r	<0.17	0.33	<0.17	0.17	<0.17	<0.17	<0.17	<0.17
Pim1	0.17	0.5	<0.17	0.5	<0.17	<0.17	<0.17	<0.17
CD83	<0.17	0.67	<0.17	0.17	<0.17	<0.17	<0.17	<0.17

Number of translocations/10^5 cells.

Table S5. Frequency of hits in genes in WT and AID$^{-/-}$ HTGTS libraries

				c-myc$^{25xI\text{-}SceI}$/WT(n=5)		c-myc$^{25xI\text{-}SceI}$/AID$^{-/-}$ (n=4)	
Chr	**Chr size (Mb)**	**Total Gene size (Mb)[a]**	**Expected rate[b]**	**Observed rate[c]**	**p value[d]**	**Observed rate[c]**	**p value[d]**
Chr1	197	70	0.36	0.41 ± 0.06	0.114	0.46 ± 0.08	0.075
Chr2	182	73	0.40	0.62 ± 0.06	0.0019	0.57 ± 0.07	0.019
Chr3	160	46	0.29	0.40 ± 0.06	0.012	0.49 ± 0.02	0.0003
Chr4	156	55	0.35	0.49 ± 0.11	0.053	0.51 ± 0.03	0.001
Chr5	153	61	0.40	0.60 ± 0.04	0.0006	0.58 ± 0.08	0.025
Chr6	150	58	0.39	0.45 ± 0.07	0.142	0.46 ± 0.07	0.144
Chr7	153	53	0.35	0.65 ± 0.05	0.0002	0.62 ± 0.05	0.001
Chr8	132	45	0.35	0.51 ± 0.09	0.014	0.49 ± 0.07	0.021
Chr9	124	51	0.41	0.52 ± 0.04	0.002	0.58 ± 0.09	0.040
Chr10	130	49	0.38	0.53 ± 0.06	0.006	0.55 ± 0.08	0.026
Chr11	122	53	0.43	0.60 ± 0.02	8.9985E-05	0.58 ± 0.01	0.0003
Chr12	121	39	0.34	0.81 ± 0.04	1.3954E-05	0.39 ± 0.02	0.012
Chr13	120	40	0.33	0.52 ± 0.08	0.006	0.48 ± 0.17	0.175
Chr14	125	42	0.34	0.54 ± 0.10	0.010	0.64 ± 0.09	0.007
Chr15	103	37	0.36	0.48 ± 0.05	0.006	0.51 ± 0.02	0.001
Chr16	98	34	0.35	0.52 ± 0.03	0.0002	0.52 ± 0.06	0.011
Chr17	95	36	0.38	0.54 ± 0.08	0.013	0.55 ± 0.06	0.012
Chr18	91	31	0.40	0.53 ± 0.08	0.024	0.39 ± 0.08	0.6194
Chr19	61	27	0.44	0.58 ± 0.04	0.001	0.71 ± 0.05	0.001
ChrX	167	42	0.25	0.36 ± 0.05	0.013	0.43 ± 0.08	0.024
ChrY	0.77	16	0.05	NA	NA	NA	NA

[a] Calculated according to refgene data available at http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/. In the case of multiple isoforms, the longest was used.
[b] Total gene size/Chr size
[c] Mean and standard deviation
[d] Calculations are with a one sample t-test

Table S6. Number of translocations to cryptic I-SceI sites (falling within a 4 kb region centered on a given site) for each HTGTS library.

	c-myc$^{25xI\text{-}SceI}$/WT								c-myc$^{25xI\text{-}SceI}$/AID$^{-/-}$					c-myc$^{1xI\text{-}SceI}$/WT			
mouse	248-249	253	311	484-485	488	749	858	**TOT**	342	409	487	494	**TOT**	2	9	183	**TOT**
chr1 (A)	1	1	2	3	2	1	0	**10**	1	6	3	0	**10**	0	2	3	**5**
chr1 (B)	3	1	5	3	3	0	0	**15**	0	6	6	3	**15**	0	1	3	**4**
chr11	0	0	1	0	1	0	0	**2**	2	2	2	0	**6**	0	1	5	**6**
chr13	0	0	0	0	0	0	0	**0**	0	0	0	0	**0**	0	1	0	**1**
chr14	2	1	8	2	4	0	0	**17**	0	7	6	4	**17**	0	2	9	**11**
chr15 (A)	1	0	0	1	1	0	0	**3**	0	1	1	0	**2**	0	0	1	**1**
chr15 (B)	2	2	12	5	5	0	0	**26**	2	11	7	1	**21**	1	5	5	**11**
chr19 (A)	0	0	0	0	0	1	0	**1**	0	2	3	3	**8**	0	1	1	**2**
chr19 (B)	2	0	0	0	1	0	0	**3**	0	0	0	0	**0**	0	3	1	**4**
chr2 (A)	0	0	1	2	1	0	0	**4**	1	0	0	0	**1**	0	1	0	**1**
chr2 (B)	1	0	0	0	1	1	0	**3**	0	4	1	2	**7**	0	0	1	**1**
chr2 (C)	2	5	12	4	13	3	2	**41**	3	19	12	6	**40**	0	5	4	**9**
chr4	0	0	1	2	1	0	0	**4**	0	1	0	0	**1**	0	1	0	**1**
chr6	0	0	0	0	1	0	0	**1**	0	2	0	0	**2**	0	1	0	**1**
chr7	1	0	5	4	4	3	0	**17**	4	3	8	2	**17**	0	2	3	**5**
chr8	0	0	1	0	1	0	0	**2**	0	1	1	0	**2**	0	0	1	**1**
chr9	0	0	5	2	1	0	0	**8**	1	8	4	0	**13**	0	4	3	**7**
chrX	1	0	3	1	1	0	0	**6**	0	0	2	0	**2**	0	0	1	**1**

	ΔSγ1$^{2xI\text{-}SceI}$/WT				ΔSγ1$^{2xI\text{-}SceI}$/AID$^{-/-}$		
mouse	186	188	958	**TOT**	165	278	**TOT**
chr1 (A)	0	0	0	**0**	0	0	**0**
chr1 (B)	0	0	0	**0**	1	0	**1**
chr11	0	0	0	**0**	0	0	**0**
chr13	0	0	0	**0**	0	0	**0**
chr14	0	0	0	**0**	0	0	**0**
chr15 (A)	0	0	0	**0**	0	0	**0**
chr15 (B)	0	0	0	**0**	0	0	**0**
chr19 (A)	0	0	0	**0**	0	0	**0**
chr19 (B)	0	3	0	**3**	0	0	**0**
chr2 (A)	0	0	0	**0**	0	0	**0**
chr2 (B)	0	0	0	**0**	1	0	**1**
chr2 (C)	1	1	2	**4**	3	5	**8**
chr4	0	0	0	**0**	0	0	**0**
chr6	0	0	0	**0**	0	0	**2**
chr7	0	0	0	**0**	0	0	**0**
chr8	0	0	0	**0**	0	0	**0**
chr9	0	0	0	**0**	0	0	**0**
chrX	0	4	0	**4**	0	1	**1**

Note that no translocation to cryptic I-SceI sites was detected in c-myc$^{25xI\text{-}SceI}$/ROSA$^{I\text{-}SceI\text{-}GR}$ libraries, likely due to lower levels of cutting at both “bait” and “prey” loci achieved with inducible I-SceI-GR as compared to retrovirally-expressed I-SceI.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Gene targeting of c-myc$^{1xI\text{-}SceI}$ and ROSA$^{I\text{-}SceI\text{-}GR}$ alleles

To generate the c-myc$^{1xI\text{-}SceI}$ targeting construct, the previously described c-myc$^{25xI\text{-}SceI}$ targeting construct (Wang et al., 2009) was digested with I-SceI and re-ligated to generate a construct containing a single I-SceI site cassette, flanked by a 4.6 kb SphI-SphI 5' homology arm encompassing *c-myc* exon1 and a 2.6 kb SphI-SphI 3' homology arm containing *c-myc* exon 2 and 3. We electroporated the targeting construct into TC1 (129/Sv) ES cells and screened EcoRI-digested DNA to identify potentially targeted ES cell clones via Southern blotting with a 5' probe consisting of the 1.5 kb XbaI-XbaI fragment upstream of *c-myc* exon 1 and a 3' probe consisting of the 1.5 kb XhoI-KpnI fragment downstream of *c-myc* exon3. To generate the ROSA$^{I\text{-}SceI\text{-}GR}$ allele we followed the previously described strategy (Sasaki et al., 2006) to target an I-SceI-GR/IRES-tdTomato expression cassette into TC1 (129/Sv) ES cells. In both cases, the Neo cassette was deleted from targeted ES clones by infection with a Cre-recombinase-expressing adenovirus before generating c-myc$^{1xI\text{-}SceI}$ or ROSA$^{I\text{-}SceI\text{-}GR}$ mice.

B cell culture conditions, retroviral infection and CSR assays

Retroviral supernatants were prepared from Phoenix packaging cells transfected with pMX-ISceI vector or empty pMX vector as control. B cells were put in culture at a density of $1x10^6$/ml in RPMI medium containing fetal bovine serum, αCD40 (1 μg/ml, eBioscience) and IL4 (20 ng/ml, R&D Systems). After 24 hours of culture, retroviral infection was performed adding one volume of viral supernatant, spinning for 1.5 hours at 2400 RPM in the presence of polybrene at 2.5 μg/ml and incubating cells with viral supernatant overnight. After infection, medium was changed and cells were replated at $0.5x10^6$/ml. At day 4 of stimulation, infection efficiency was evaluated by measuring the percentage of cells expressing the retroviral IRES-GFP and was found to be usually between 50% and 85%. ROSA$^{I\text{-}SceI\text{-}GR}$ B cells were cultured in RPMI medium as above but with 15% charcoal-stripped fetal bovine serum to minimize nonspecific activation of the I-SceI-GR fusion and thus obtain high levels of cutting at the appropriate time. CSR was evaluated in the case of retrovirally infected cells by staining with

Cy5-PE-labeled anti-mouse B220 (eBiosciences) and PE-labeled anti-mouse IgG1 (BD Biosciences). The TA-induced ROSA$^{I\text{-}SceI\text{-}GR}$ B cells were evaluated by staining with Cy5-PE-labeled anti-mouse B220 and FITC-labeled anti-mouse IgG1 (BD Biosciences). CSR ranged between 15% and 40% for retrovirus-infected B cells and 9-12% for TA-induced B cells, due to lower proliferation rates in charcoal stripped serum. DNA was prepared from cells at day 4 of culture by standard methods.

Two color FISH

Metaphases were prepared from day 4-stimulated c-myc$^{25x\ I\text{-}SceI}$ and c-myc$^{1x\ I\text{-}SceI}$ B cells infected with either control or I-SceI-expressing retrovirus, following standard protocols (Wang et al., 2009). FISH was performed with the following BAC probes flanking the *c-myc* locus: RP24-434C10 (centromeric to c-myc) and RP23-113O21 (telomeric to c-myc). The intact *c-myc* locus shows colocalization of signals from the two probes, whereas c-myc breaks and translocations are visualized as split probe signals.

Generation of HTGTS libraries by adapter-PCR

Genomic DNA from c-myc$^{25x\ I\text{-}SceI}$, c-myc$^{1x\ I\text{-}SceI}$ and c-myc$^{25x\ I\text{-}SceI}$/ROSA$^{I\text{-}SceI\text{-}GR}$ B cells was digested overnight with HaeIII and HaeIII-generated blunt ends were A-tailed with Klenow polymerase. Genomic DNA from $\Delta S\gamma1^{2x\ I\text{-}SceI}$ B cells was digested overnight with MspI. An asymmetric adapter (composed of an upper linker and a lower 3'-modified linker) was then ligated to fragmented DNA at a molar ratio of 20:1 for 30 minutes at 25 ºC. To remove unrearranged I-SceI cassettes, ligation reactions were digested either with I-SceI, or with both EcoRV and XbaI (for c-myc libraries). The blocking digests were carried out for 8 hours. Translocation junctions were then PCR-amplified using the emulsion (em)-PCR approach as previously described (Williams et al., 2006). In the first round of PCR, 1 µg of DNA was amplified in a final volume of 50 µl, using a biotinylated forward primer (Myc-L for c-myc libraries or Sγ1-C for ΔSγ1 libraries; see table below for primer sequences) and an adapter-specific reverse primer (AP1) and Phusion polymerase (Finnzymes). 20 PCR cycles were performed in the following conditions: 98ºC for 10 seconds, 58ºC for 30 seconds, and 72ºC for 30 seconds.

Multiple reactions were performed in generating large-scale libraries. Thereafter, biotinylated PCR products were isolated using the Dynabeads MyOne Streptavidin C1 kit (Invitrogen) and an additional 3-hour digestion with blocking enzymes was performed. PCR products were eluted from the beads via a 30 minute incubation at 65ºC in 95% formamide/10 mM EDTA and purified. The purified products were then amplified in a second round with em-PCR. 30 µl of the first round PCR product, 80 pmol each of primers AP2 and Myc-Prelox (for c-myc libraries) or Sγ1-E (for ΔSγ1 libraries), 20 µl ultra-pure BSA (Ambion) and 2 µl Taq polymerase (Qiagen) in a final volume of 200 µl were emulsified in 400 µl of an oil-surfactant mixture. The emulsion mixture was divided into 50 µl individual aliquots and PCR was performed using the following conditions: 30 cycles of 94ºC for 30 seconds, 60ºC for 30 seconds, and 72ºC for 1 minute. Following PCR, the products were pooled and centrifuged in a table-top centrifuge for 5 minutes at 14,000 RPM to separate the phases and the oil layer was removed. The sample was then extracted 3 times with 1 ml of H_2O-saturated diethyl ether and DNA was re-purified. The third, non-emulsion, round of PCR was performed with the same primers as in round 2, but with the addition of linkers and barcodes for 454 sequencing. After amplification, the pooled PCR reactions were size-fractionated between 200 and 800 base pairs on agarose gel. An aliquot of the gel-purified material was submitted for 454-sequencing.

Generation of HTGTS libraries by circularization-PCR

Genomic DNA from c-myc$^{25xI\text{-}SceI}$, c-myc$^{1xI\text{-}SceI}$ and c-myc$^{25xI\text{-}SceI}$/ROSA$^{I\text{-}SceI\text{-}GR}$ B cells was digested overnight with HaeIII. Genomic DNA from ΔSγ1$^{2xI\text{-}SceI}$ B cells was digested overnight with MspI. After purification 250 ng of DNA was ligated for 12-hours in 150 µl total volume to favor formation of intramolecular circles. To obtain HTGTS libraries, multiple ligations were performed for each sample and the final purified ligation products were pooled together. The ligated material was heat inactivated at 65ºC for 30 minutes and then incubated with EcoRV and XbaI (for c-myc libraries) or EcoRI, HindIII, and PstI (for ΔSγ1 libraries) for 4 hours. These blocking digests served to linearize circular products deriving from WT (non-targeted) alleles and from unrearranged I-SceI cassettes. The DNA was then purified and resuspended in a final

volume of 50 µl. Nested PCR was either performed in multiple reactions, using 1 µl of purified ligation mixture each (which yielded on average at least one junction per reaction) or in bulk with the em-PCR approach detailed above. The first PCR amplification step was carried out with Myc-G and Myc-H primers for c-myc libraries, or Sγ1-B and Sγ1-C for ΔSγ1 libraries (see table for primer sequence) and Phusion polymerase (Finnzymes). 20 PCR cycles (for em-PCR) or 30 PCR cycles (for separate reactions) were performed as follows: 98ºC for 10 seconds, 58ºC for 30 seconds, and 72ºC for 30 seconds. For large-scale libraries, 60 em-PCR reactions of 50 µl final volume were run, using 500 ng of ligated DNA per each reaction. For the second PCR amplification step we used primers Myc-I and Myc-Prelox (for c-myc libraries) or primers Sγ1-A and Sγ1-E (for ΔSγ1 libraries) and Taq polymerase (Qiagen). 30 cycles of PCR amplification were performed with the following conditions: 94ºC for 30 seconds, 60ºC for 30 seconds, and 72ºC for 1 minute. For the multiple, non-emulsion PCR reactions, round 2 PCR primers contained the linker fragment and barcode as required for 454 sequencing. In the case of em-PCR, these linkers and barcodes were added in a third round of PCR (10 cycles). After amplification, PCR products were pooled and 200-1000 bp fragments were isolated after fractionation on agarose gel and submitted for 454-sequencing.

Calculation of restriction sites distribution

We defined all recognition sequences for HaeIII or MspI in the mouse genome and calculated the distances between two consecutive sites. The distribution of the distance showed that the probability of getting a fragment of more than 10 kb for each of the cutting enzyme was less than 0.002% for HaeIII and 1.72% for MspI. Such distribution should allow cloning of 90% or more of potential junctions by out HTGTS approaches.

Data analysis

Alignment and filtering. Raw 454 sequence data was aligned to the mouse reference genome (Mouse July 2007-NCBI Build37/mm9) with the BLAT program (Kent, 2002). Data were then filtered to remove PCR repeats (including repeats that are slightly divergent due to sequencing errors), fragment ligation artifacts (where, for example, a random HaeIII or MspI

fragment was incorporated during *in vitro* ligation), 'illegitimate' ligation products (rare events in which a fragment generated by the HaeIII or MspI became ligated *in vitro* to an *in vivo* generated I-SceI end), and mis-priming during PCR. Specific settings for BLAT and scripts for the filters are available on request. For hotspot analysis, candidate hits were confirmed by repeating the alignment using the NCBI BLAST webserver (default settings) (Johnson et al., 2008) and by locating and inspecting the sequence spanning the translocation junction. Hits with discrepancies between BLAST and BLAT alignment and bearing junctions that had already been scored in the same mouse/DNA sample were removed. These post-filtering steps were simply executed and were critical for hotspot designations to prevent infrequent artifacts from confounding the analyses.

Data display. To visualize the distribution of junctions within HTGTS libraries circle plots were generated using Circos program (Krzywinski et al., 2009). As input data for Circos, the coordinates of the breaksite and the translocation junctions were used. To generate translocation maps, we also used R program (Development Core Team, 2010) to calculate the position of each individual junction relative to the chromosomal length (divided into bins of different size as specified in Figure legends) and then generated a dot to represent a single translocation junction. For comparison of translocation density maps and nascent RNA signals, we uploaded both data sets to the UCSC genome browser.

Determination of translocation hotspots. All filtered reads derived from WT or $AID^{-/-}$ libraries were pooled for hotspot analysis. Reads aligning to chr15 or to the *IgH* locus were eliminated from each pool. The adjusted genome was divided into equal sized bins (phased with the first nucleotide of each chromosome), and bins occupied with zero, one, two, etc. hits were enumerated. The bin size was adjusted down from 2 Mb until a size for which 60% of the bins were empty. The data fit a Poisson distribution. Hotspots were defined as all bins in the 10% of the tail of the distribution ($P< 0.05$). For both the WT and $AID^{-/-}$ pools, we used a bin size of 250 kb. Bins containing ≥ 5 hits constituted a hotspot. When pools were re-analyzed after offsetting bins by 190 kb hotspots were similar in number and location.

Evaluation of HTGTS background.

Human DNA isolated from the K562 cell line was mixed 1:1 with DNA from αCD40/IL4-activated and I-SceI-infected c-myc$^{25xI\text{-}SceI}$ or ΔSγ1$^{2xI\text{-}SceI}$ primary B cells from WT or AID$^{-/-}$ background. The DNA mixes were then used to generate HTGTS libraries as described. The primary sequence readouts were analyzed against the mouse genome to identify the real translocations as detailed above. In addition, the same primary readouts were analyzed against the human genome with the same filter settings, to identify potential artifactual translocations involving human DNAs. Alignments were verified by BLAST. The background was calculated as percentage of artifactual human:mouse hybrid junctions over total junctions.

Analysis of translocation junctions.

Small-scale HTGTS libraries obtained from one c-myc$^{25xI\text{-}SceI}$ (1359 junctions), one c-myc$^{1xI\text{-}SceI}$ (1066 junctions) and one c-myc$^{25xI\text{-}SceI}$/ROSA$^{I\text{-}SceI\text{-}GR}$ mouse (547 junctions) were extensively and individually post-filtered to remove all duplicate sequences (defined as identical junctions) and ambiguous alignments. MH at the junction was calculated with two different algorithms, and sequences that differed were further inspected. Junctions mapping within 1 Mb from the breaksite (chr15: 61818880) on either side were considered separately. Junctions corresponding to local resection events (within the segment defined by the first HaeIII site telomeric to the I-SceI cassette) were not included in the analysis.

Identification of cryptic I-SceI sites.

Candidate I-SceI sites were identified by examining the sequences of all recurrent hits in the WT or AID$^{-/-}$ library pools displaying an exceptionally focal pattern. In addition, all translocation junctions were compared against a hit table of all sites in the mouse genome (± 200 bp) matching the canonical I-SceI consensus sequence at 15 or more positions. A genomic region encompassing each candidate I-SceI site then was PCR-amplified with primers detailed below. This approach was taken to avoid potential effects of Dam and Dcm methylation on cutting efficiency (Petek et al., 2010). The genomic sequence of the candidate I-SceI sites was confirmed by sequencing the corresponding PCR product. 500 ng of purified PCR products were digested *in vitro* with 5 units of I-SceI for 3 hours. Digestion

reactions were separated on agarose gel and the relative intensity of the uncut and I-SceI-digested bands were calculated with the FluorchemSP program (Alpha Innotech). Efficiency of cutting was scored as follows: +++, 100% to 70%; ++, 70% to 30%; +, 30% to 10%; -, less than 10%.

PCR assay to detect translocations between c-myc$^{25x\ I\text{-}SceI}$ and hotspot genes

Translocation junctions were PCR-amplified from DNA prepared from αCD40/IL4-activated c-myc$^{25xI\text{-}SceI}$ B cells infected with control or I-SceI-expressing retrovirus. To estimate translocation frequencies, multiple independent PCR reactions were performed each containing 5000 or 50000 cell equivalents of DNA. Two rounds of nested PCR reactions were performed. Reverse primers Myc-Ex2b and Myc-Ex2a (see table for primer sequence) were specific for *c-myc* sequence telomeric to the I-SceI cassette. Forward primers were specific for each gene tested (*Dmrt1, Scd2, Mmp24, Il4ra, Il21r, CD83* and *Pim1*) and are listed in the table below. For cryptic I-SceI targets, these primers were chosen centromeric to the putative site. For the first PCR round, conditions were as follows: 94°C 3min; 94°C 15s, 62°C 15s, 68°C 7min plus an increment of 20s per cycle, 25 cycles; 68°C 5 min. For the nested PCR round, conditions were as follows: 94°C 3min; 94°C 15s, 62°C 15s, 68°C 7min, 25 cycles; 68°C 5 min. The Expand Long Template PCR System (Roche) was used for both reactions.

Analysis of GRO-seq data

We calculated promoter proximal gene activity indexes at transcriptional start sites (TSSs) by analyzing the GRO-seq density ±1 kb around the TSSs as defined for 21,906 genes annotated in the current reference sequence of the mouse genome (NCBI Build 37, mm9) from the UCSC server (http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/refGene.txt.gz). Where a gene had multiple isoforms, the longest was used. Each 2 kb window was divided into 200 bp bins and tiled across in 50 bp increments to identify the peak GRO-seq signal. We obtained a promoter proximal gene activity and a p value based on our background density, which we calculated to be 0.02 reads/kb. If the promoter proximal peak had a p value less then 0.001, we considered the gene to have significant promoter proximal activity and therefore such

a gene was called “active”, if had non-significant promoter proximal peak activity then the gene was called “inactive”. For calculation of distance of a translocation to the nearest TSS, we used the coordinate of the reference sequence defined by refGene.

RefGene annotated TSSs and Gro-Seq determined promoter proximal peaks were quite close. For active genes, if we utilized the promoter proximal peak coordinate instead of the TSS based on reference sequence we found that the overall pattern of proximity to transcription was similar, but that translocations tended to fall closer to promoter proximal peak coordinates as compared to the refGENE defined TSS.

Table of oligonucleotides used in this study

NAME (orientation)	SEQUENCE	PURPOSE
c-myc primers		
Myc-Prelox (For)	ACCGCCGCTAATTCCGATCATATTC	Amplifying 25x cassette; 2nd round PCR for HTGTS (ad-PCR and cir-PCR)
Myc-Ex2a (Rev)	ATAGGGCTGTACGGAGTCGTAGTC	Amplifying 25x cassette; 2nd round direct translocation PCR
Myc-Ex2b (Rev)	GCTCTGCTGTTGCTGGTGATAGAA	1st round direct translocation PCR
Myc-G (Rev)	CCTTCGAGCAGGGACTTAGCC	1st round PCR for HTGTS (cir-PCR)
Myc-H (For)	AGCAGCTGCTAGTCCGACGA	1st round PCR for HTGTS (cir-PCR)
Myc-I (Rev)	AGACGCCCAGGAATCGCCAT	2nd round PCR for HTGTS (cir-PCR)
Myc-L (For)	CGAGCGTCACTGATAGTAGGGAGT	1st round PCR for HTGTS (ad-PCR)
Sγ1 primers		
Sγ1-B (Rev)	GCTCAGGTTTGTCTGTGGG	1st round PCR for HTGTS (cir-PCR)
Sγ1-C (For)	GGAATATATCGAGAAGCCTGAGG	1st round PCR for HTGTS (ad-PCR and cir-PCR)
Sγ1-A (Rev)	CTC TAC ATG CCT GTG CTT GT	2nd round PCR for HTGTS (cir-PCR)
Sγ1-E (For)	GCCTCGAGGGACCTAATAAC	2nd round PCR for HTGTS (cir-PCR and ad-PCR)
Other primers for ad-PCR		
HaeIII upper linker	GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGTTAT	
HaeIII lower linker	TAACCAGCCC (5'P, 3'inverted dT)	
MspI upper linker	GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGT	
MspI lower linker	CGACCAGCCC (5'P, 3'inverted dT)	
AP1	GTAATACGACTCACTATAGGGC	
AP2	ACTATAGGGCACGCGTGGT	
Primers for amplification of genomic region containing cryptic I-SceI sites		
chr1A-for	CCCTGCCTCCCTTAAGCAGGC	
chr1A-rev	CAGACTTGGAGGGAGGGAGGG	
chr2A-for	TCAAAGATGTTTGGAGGCCACTGC	
chr2A-rev	GCTCATGCAACACTACCCTGTGC	
chr2B-for	TCTGAGAGTCTTCCCAGTCCACC	
chr2B-rev	GGGAAGTGCACAGCAAACCGG	
chr2C (Mmp24)-for	AATCCTGGTACTGGAATCGG	
chr2C (Mmp24)-rev	AGATGGCACACTGCTAGGAC	
chr4-for	ACTGCTCTGCCTGGTCACCTTG	
chr4-rev	TGCACGTGATTCCAGCACTTGG	
chr7 (Apbb)-for	TCCACACTCACTACAGCCATTTCC	
chr7 (Apbb)-rev	TCTCTTGTTCCCCTCATACCTCAC	
chr8 (PaIId)-for	GACACACTGCAAAAGCTGATCCCC	
chr8 (PaIId)-for	TCCTGCCAGTCACCTCCCATC	
chr9 (kirrel3)-for	ACAAGGCACCACACAATGGGTG	
chr9 (kirrel3)-rev	TTAAGGGCACTGTCTGAGGGGG	
chr13-for	AACCCTCATACTCTCCAGCCACC	
chr13-rev	TTGCTGCCCTATGAGTCCTGGG	
chr14 (Fermt2)-for	GAGGCTCACCTTTCAGAGGGCC	
chr14 (Fermt2)-rev	GAGGCTTCCCTTATCGTGGCACC	
chr15A-for	GGCATCGCATCCAAGTCAACCTC	
chr15A-rev	CCTTTTGAGGGCCACCTGACTG	
chr15B (Aco2)-for	TTAAGGTGGTCTTCCCCTGAGGC	
chr15B (Aco2)-rev	AAGAGGCTGTCAGTGAGCCGAG	

chrX-for	AGTGAGAAATGAGCACCATGGGC	
chrX -rev	GAGGCACGAAGAGGCTACTCAG	
Primers for direct tranlsocation PCR		
CD83-ExtF	TGCTTACGCCGCTCTGTTTCT	First round
CD83-IntF	TATGCAGTGTCCTGGCCAAG	Second round
Dmrt1-ExtF	GACCTACCACTCTGCAGCTGG	First round
Dmrt1-IntF	GGAGTTCTAGGCCAGCCTTGG	Second round
Il21r-ExtF	ATGTCCTCCTTCCCACAATGCTG	First round
Il21r-IntF	AGCAGTGCTTAAGGCAGAAAGTCTG	Second round
Il4ra-ExtF	GCCTGAACTCGACGGTAGGAAC	First round
Il4ra-IntF	AGAACCGATCTGGCCTGAAACC	Second round
Mmp24-ExtF	GGTCACTAACTCATGCCCCACC	First round
Mmp24-IntF	GAGGAGACGGAAGTGAAGCTCTG	Second round
Pim1-ExtF	TGGCCATTAAGCACGTGGAGAAG	First round
Pim1-IntF	GATTGGGGAGAACTGGTGAGTGAG	Second round
Scd2-ExtF	AAGGGCACAAGTTAGGTGGTAGGA	First round
Scd2-IntF	AGGGTGTGAGAGGAAAATGGTGG	Second round

SUPPLEMENTAL REFERENCES

Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Res *36*, W5-9.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome Res *12*, 656-664.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. Genome Res *19*, 1639-1645.

Petek, L.M., Russell, D.W., and Miller, D.G. (2010). Frequent endonuclease cleavage at off-target locations in vivo. Molecular therapy *18*, 983-986.

Sasaki, Y., Derudder, E., Hobeika, E., Pelanda, R., Reth, M., Rajewsky, K. and Schmidt-Supprian, M. (2006). Canonical NF-kappaB activity, dispensable for B cell development, replaces BAFF-receptor signals and promotes B cell proliferation upon activation. Immunity *24*, 729-739.

Williams, R., Peisajovich, S.G., Miller, O.J., Magdassi, S., Tawfik, D.S., and Griffiths, A.D. (2006). Amplification of complex gene libraries by emulsion PCR. Nat Methods *3*, 545-550.