FIGURE LEGENDS FOR SUPPLEMENTARY FIGURES

Supplementary Figure 1. Genome-wide distribution of translocations from G1-arrested A-MuLV transformed WT and ATM-/- pro-B cells.

A) Translocations originating from the I-SceI cassette in chr15, labeled by blue arrow, from G1-arrested A-MuLV transformed WT pro-B cells. The genome was divided into 2Mb bins and the number of unique translocations within each bin is represented by colored dots with a black dot indicate 1 translocation, a red dot 5, a yellow dot 20 and a green dot 100, respectively. The junctions from STI571 and TA treated cells are plotted on the left side of each chromosome ideogram, while translocations from STI571, TA, and ATMi-treated cells are plotted on the right side. Ig/TCR hotspots are indicated by the green arrows. Centromere (Cen) and telomere (Tel) positions are indicated. Data are from pooled HTGTS libraries. **B)** Background for HTGTS approaches, calculated as percent of artifactual human:mouse hybrid junctions when human DNA was mixed 1:1 with mouse DNA from indicated samples.

Supplementary Figure 2. Genome-wide distribution of translocations from Chr7 I-SceI breaksite .

Genome-wide map of translocations originating from the I-SceI cassette in chr7 (labeled by blue arrow). Other details are the same as for Figure S1A.

Supplementary Figure 3. Genome-wide and locus specific distributions of translocations and the association of breaksite with Igκ.

A) Translocations from chr18$^{I\text{-}SceI}$ in the 3.2 Mb IgK region are represented. (+) and (-)-oriented junctions are plotted on top and bottom of chromosome diagrams, respectively. **B)** The distribution of translocations around the integrated I-SceI substrate in chr18. 10kb and 2.5Mb regions were presented. (+) and (-)-oriented junctions are plotted on top and bottom of chromosome diagrams, respectively. Sequencing primer for HTGTS is indicated. **C)** All three integration sites show interactions with Igκ that have Hi-C scores higher than the median value of all other *trans* interactions with Igκ. The histogram shows the distribution of Hi-C interaction scores between all 1 Mb *trans* bins and Igκ. The vertical lines show the Hi-C interaction score

between 1 Mb around each I-Sce site and Igκ. **D)-F)** Whole chromosome distribution of translocations. Log ratios of observed/expected translocation frequencies for each chromosome are calculated for each replicate separately, then averaged and displayed with ± SEM. Results are shown for chr18$^{I\text{-}SceI}$ (**D**), chr2$^{I\text{-}SceI}$ (**E**), and chr7$^{I\text{-}SceI}$ (**F**) upon IR treatment. Red arrows indicate breaksite chromosomes. Chromosomes are sorted from longest to shortest.

Supplementary Figure 4. The distribution of allele specific translocations in F1 cells.

A, B) Genome-wide map of translocations originating from the DEL-CJ cassette, labeled by blue bar, for chr7$^{DEL\text{-}CJ}$ (**A**) and chr9$^{DEL\text{-}CJ}$ (**B**), from A-MuLV transformed F1 mouse pro-B cells. The genome was divided into 2Mb bins and the number of unique translocations within each bin is represented by colored dots with a black dot indicate one translocation, a red dot 5, a yellow dot 20 and a green dot 100, respectively. **C, D)** Allelic specific translocation junctions for chr7$^{DEL\text{-}CJ}$ (**C**) and chr9$^{DEL\text{-}CJ}$ (**D**) from F1 cells. Single junctions from BALB/cJ alleles are plotted on left side of each ideogram, while translocations from 129 alleles are plotted on right side. **E)-J)** allelic specific translocations from individual HTGTS libraries are represented in circos plots. Individual translocations are represented as arcs originated from specific I-SceI breaks, indicated by arrows, and terminating at partner site.

Supplementary Figure 5. Similarity of Hi-C replicates and WT and ATM$^{-/-}$ Hi-C data.

A) Heat maps showing interactions between all regions along mouse chr18 demonstrate the high reproducibility of the Hi-C data across biological replicates (R1 vs. R2) and cell lines with different I-SceI integration sites. **B)** The same chr18 heat maps for WT A-MuLV G1 arrested pro-B lines show the similarity of spatial genome organization between ATM$^{-/-}$ cells (A) and WT cells. **C)-G)** The same analyses (all by all heat map (C), chr18 heat map (D), correlation and compartment analysis (E), scaling analysis (F), and whole chromosome positioning analysis(G)) are performed for WT cells as were performed for ATM -/- cells in Fig 5. See Fig. 5 legend for details. Compartment eigenvector values in **E)** are shown for both chr18 (top) and chr2 (bottom) with a direct comparison between the WT compartment eigenvector and the

$ATM^{-/-}$ eigenvector. The overall compartment structure is very similar between WT and $ATM^{-/-}$, but does show some differences (see Text).

Supplementary Figure 6. Hi-C resolution and *cis* correlations for chr7

A) Hi-C data coverage as a function of fixed and partner bin size. Each graph shows the percentage of *cis* (top) or *trans* (bottom) bin pairs that contain at least one mapped read from a valid interaction pair (y-axis) for each different partner bin size (x-axis). The effect of different fixed bin sizes (anchored on the I-SceI site) is shown in different colors. At the sequencing depth of this experiment (~184 million reads), 100% of interaction pairs in *cis* are represented by at least one sequencing read when a 1 Mb fixed bin size is used, even with a partner bin size as small as 100 kb. With a more focused fixed bin (100 kb), a 1 Mb partner bin size is required to achieve 100% coverage of bins in *cis.* In *trans*, at least a 500 kb partner bin size is required to achieve 100% coverage with a 1 Mb fixed bin size while a 5 Mb partner bin size is required for a 100 kb fixed bin size. **B)** The minimum resolution of the translocation data is somewhat lower than that of the Hi-C data. In *cis*, at least one junction occurred within each bin for a bin size of 500 kb, while in *trans*, a bin size of 5 Mb was required for at least one translocation junction to mapped within each bin. Bins with DNA sequence that is too repetitive to allow sequence read mapping are excluded from each of the analyses in A) and B). **C)** The correlation in *cis* between Hi-C interactions and translocation frequency changes as a function of the size of the fixed bin around the I-SceI site. Each correlation is calculated using a partner bin size of 1 Mb for Hi-C and translocations. For each different I-SceI site (chr2, chr7, and chr18 integrations), the correlation is lower for a small fixed bin size (100 kb) than for a somewhat larger fixed bin size (500 kb – 1 Mb). Since the combination of a 100 kb fixed bin with a 1 Mb partner bin gives 100% coverage of all bins (see A) and B)), this lower correlation is not likely due to a decrease in statistical power. **D)** Hi-C and translocation data as in Fig. 6, anchored on the chr7 I-SceI site. **E)** Scatter plots for log (IR translocations) vs. log(Hi-C) for the chr7 I-SceI integration as in Fig. 6.

Supplementary Figure 7. Additional trans Hi-C vs. translocation analyses

A) The difference in Hi-C score between high and low translocation Hi-C scores (Fig 7A) is significant across a broad range of translocation thresholds. Each point in the panels represents the negative log10 p-value for the one-sided KS test testing the difference between cumulative curves of Hi-C scores for "high" and "low" translocation bins as in Fig 7A. The threshold (in units of translocations/1,000 in whole dataset) above which a bin is classified as having "high" translocations is shown on the x-axis. A permutation test threshold above which the difference between distributions is significant is indicated by a dotted line on each plot. **B)** Genomic regions that frequently form interchromosomal translocations with I-SceI sites (blue; >=2 translocations/1,000 in dataset) have higher Hi-C interaction scores with those I-SceI sites than regions that do not translocate frequently (red; < 2 translocations/1,000 in dataset). Hi-C scores between a 1 Mb bin fixed at the I-SceI site and 5 Mb trans bins are averaged over the set of high translocation trans bins (blue) or low translocation trans bins (red). The Hi-C scores across 20 Mb upstream and downstream of each high or low translocation bin are aligned to center on the high or low translocation bin and then averaged to give the signatures shown. Randomly selected bins with the same sample size as the high and low translocation bins give Hi-C score profiles that are indistinguishable from one another. All Ig and TCR hotspot loci are excluded from this analysis. **C)** Short chromosomes (chr2) and long chromosomes (chr18) have a stronger and more reproducible whole chromosome Hi-C pattern than mid-length chromosomes (chr7). Obs/Exp values are calculated as in Fig 5E for each of the 5 Hi-C replicates shown in Fig S5A. Error bars show 1 standard deviation of the mean of these 5 Hi-C replicates. **D)** Cumulative Hi-C score distribution for high and low translocations between chr2 I-SceI and chr9. The whole chromosome translocation and interaction frequency between chr2 I-SceI and chr9 are discordant, as shown in Fig 7B. However, interactions between a 5 Mb partner bin along chr9 and a 1 Mb fixed bin around the chr2 I-SceI site are higher for bins with more frequent translocations. **E)** For the chr7 I-SceI integration, the nuclear position of a 1 Mb

fixed bin (right) around the breaksite shows a higher correlation with translocation frequency than the whole chromosome positioning (left). Obs/exp values are calculated as in Fig 5E.

EXTENDED EXPERIMENTAL PROCEDURES

High Throughput Genomic Translocation Sequencing (HTGTS)

A-MuLV transformed pro-B cell lines were arrested in G1 by STI571 treatment for 4 days while TA was added to induce I-SceI-GR activity. In some experiments, cells were treated with 5Gy IR at day2. Genomic DNA was purified and digested with MseI enzyme. Digested DNA was ligated at 1ng/µl overnight by T4 DNA ligase. The ligated product was purified and digested with NotI and I-SceI enzymes to linerize the PCR templates and remove unrearranged I-SceI cassette. Three rounds of PCR were performed as described (Chiarle et al., 2011). PCR amplicons were sequenced by Roche 454 platform. Raw sequences were analyzed and mapped as describe (Chiarle et al., 2011).

Genome Wide Chromosome Conformation Capture: Hi-C

A-MuLV transformed pro-B cell lines were arrested in G1 by STI571 treatment for 2 days and then crosslinked in 1% formaldehyde. 25 million cells were used for each Hi-C experiment. Cells were lysed and chromatin was digested with HindIII as described previously (Lieberman-Aiden, et al., 2009). Digested ends were filled in with biotinylated dCTP and then ligated for 4 hours at 16 C. After reversing the formaldehyde crosslinks by incubation at 65 C with Proteinase K overnight, unligated biotinylated ends were removed by 4 hours of incubation with 15 U T4 DNA polymerase at 20 C in the presence of a low concentration of nucleotides (25 µM each dNTP). This exonuclease step was found to be essential to isolate ligated interaction products from unligated biotinylated ends. The DNA was fragmented by Covaris sonication to an average size of 200 bp and then the ideal size for Illumina sequencing (100-300 bp) was selected by Ampure fractionation (0.9x Ampure beads to remove the DNA fragments >300 bp followed by 1.3x Ampure beads to isolate the remaining DNA fragments > 100 bp). The DNA ends were prepared for Illumina sequencing adapter ligation by repairing the DNA ends and adding an ‘A’ to each end. The biotinylated junctions were then pulled down using MyOne

streptavidin beads at a ratio of 1 uL beads per ng of biotinylated DNA. Illumina paired end adapters were ligated onto the DNA ends and then the fragments were PCR amplified for the minimum number of cycles necessary to generate 10 nM final DNA concentration. Samples were sequenced on an Illumina HiSeq instrument using the Paired End 50 bp module so that 50-100 million 50 bp paired-end reads were obtained per sample.

Hi-C Sequence Mapping

Sequencing reads from the Hi-C experiment were mapped to all possible HindIII fragments in the mouse genome (mm9) using NovoAlign V2.07.11 with default parameters. If a sequence read contained the junction between two ligated HindIII sites (indicating that the read had passed into the paired interacting fragment), only the sequence upstream of this recognized junction was mapped. Paired end reads that mapped to the same restriction fragment (representing self-ligated or unligated sequences) were discarded as were redundant observations of an identical mapped pair of locations. Such redundant read pairs likely arise from a single interacting fragment pair that was amplified by PCR.

Data Normalization and Binning

For each different analysis, Hi-C interaction counts were binned at either 250 kb, 1 Mb, or 5 Mb. Restriction fragments longer than 100 kb were excluded as they might behave unexpectedly in the experiment or represent regions in which restriction sites were not correctly annotated in the genome due to repetitiveness. Genomic bins in which > 50% of the sequence length was excluded due to large restriction fragments or unmappable sequence were marked as “NA” and excluded from analyses. Different genomic bins may have different biases in terms of mappability, number and length of restriction fragments, GC content, etc. (Yaffe and Tanay 2011). Therefore, we correct the Hi-C interaction map according to the coverage of each bin in the experiment. This coverage correction does not require knowledge of specific biases, but uses the genome-wide raw data to account for any systematic biases of particular bins in the dataset. The correction divides the number of interactions between two bins by the product of

the number of reads ever observed in each bin across the genome, using the following equation:

$$\text{Hi-C score for Bin1 \& Bin2 interaction} = \left(\frac{\#\ \text{Bin 1 and Bin2 Interaction Reads}}{\left(\frac{\text{Total Reads from Bin 1}}{\text{Total Reads in Dataset}}\right) * \left(\frac{\text{Total Reads from Bin 2}}{\text{Total Reads in Dataset}}\right) * \text{Total Reads in Dataset}} \right)$$

This correction will substantially reduce the variation in the total sum of Hi-C scores between each genomic bin and the rest of the genome. However, a complete correction requires iteratively repeating this calculation on each successively corrected heatmap. For these datasets, 10 iterations is sufficient to correct systematic biases in the dataset. We directly compared our correction approach with the computationally more intense method proposed by Yaffe and Tanay (Yaffe and Tanay 2011) and found that the methods yield equivalent results.

To compare Hi-C and translocation data, we binned the counts from each dataset and then smoothed the data using a 200 kb (for 1 Mb binning) or 1 Mb (for 5 Mb binning) step size. Translocations from different I-SceI integration sites were normalized to the same scale by dividing the number of translocations in each bin by the total number of translocations in the dataset, multiplied by 1,000. (Thus, each translocation measurement is represented as a number of translocations per 1,000 translocations in the dataset).

Compartment Analysis

Compartments (Fig. 5C and S5E) were identified as previously described (Lieberman-Aiden et al., 2009). Briefly, the expected number of Hi-C reads between bins separated by each genomic distance was calculated using a loess smoothed average over the dataset. The log ratio of observed Hi-C reads to this expected value was then calculated. The Pearson correlation between the pattern of chromosomal interactions at each pair of bins was then calculated, and this correlation matrix was used to perform Principle Components Analysis. The eigenvalue of the first principle component was then plotted as the compartment assignment, with positive values corresponding to high gene density (“compartment A” or “open chromatin”) and negative values corresponding to low gene density (“compartment B” or “closed

chromatin"). The gene density was determined by calculating the number of genes in each bin according to the UCSC KnownCanonical table of mouse genes.

Hi-C and Translocation Resolution Evaluation

The percentage of bins containing at least one Hi-C sequencing read or translocation junction was chosen as a measure of dataset resolution (Fig S6). For Hi-C data, different fixed bin (around the I-SceI site) and partner bin sizes were tested and the percentage of partner bins containing at least one interaction read was measured. If 100% of bins contain at least one read, then each bin can be evaluated, at least to a certain extent, for its proximity to the I-SceI site. *Cis* and *trans* data were considered separately because each was expected to have a different resolution.

Interchromosomal Interactions and Translocations Analyses

5 Mb bins on all trans chromosomes were classified as having "high translocations" (blue) or "low translocations" (red) with 1 Mb around the I-SceI site based on a threshold (multiple thresholds were tested). The cumulative distribution of Hi-C scores in each category was calculated and a one-tailed KS test was used to evaluate whether the "high translocation" bins had significantly higher Hi-C scores than "low translocation" bins. To test the significance of this separation, raw translocation data were permuted 1,000 times and then rebinned and smoothed. These permuted translocations were then used to calculate a background distribution of KS test p-values comparing randomly selected sets as "high" and "low" translocations. With this background distribution, we found the true level of significance of the original result. The average profile of Hi-C scores around high or low translocation bins were also calculated. Average Hi-C scores with a 1 Mb bin fixed at the I-SceI site were calculated for each 5 Mb bin for 20 Mb upstream and downstream of bins with frequent or infrequent translocations. All hotspot loci were excluded from these analyses.

Whole Chromosome Positioning Analysis

The relative frequencies of interactions between each whole chromosome in the mouse genome was determined by comparing the observed number of Hi-C reads involving

interactions between two chromosomes to the number of interactions expected to occur between them according to the total numbers of reads observed on each chromosome. This calculation was performed as follows:

$$\left(\frac{\#\ \text{Chr1 \& Chr2}}{\left(\left(\frac{\text{Chr1 Trans}}{\text{Total Trans}}\right)\left(\frac{\text{Chr2 Trans}}{\text{Total Trans} - \text{Chr1 Trans}}\right) + \left(\frac{\text{Chr2 Trans}}{\text{Total Trans}}\right)\left(\frac{\text{Chr1 Trans}}{\text{Total Trans - Chr2 Trans}}\right)\right) * \left(\frac{Total\ Trans}{2}\right)}\right)$$

where “Chr1 Trans” is the number of interactions between Chr1 and other chromosomes and “Total Trans” is the total number of interchromosomal interactions in the dataset. The two terms in the denominator represent the probability of choosing a read from chr1 first and then choosing chr2 from the remaining pool of trans reads that do not include chr1 OR (‘+’ sign) choosing a read from chr2 first and then choosing chr1 from the remaining pool of trans reads that do not include chr2. The observed number of translocations between the breaksite and each chromosome was similarly compared to the expected number of translocations according to the length of each chromosome. We excluded known hotspots of RAG breaks (Ig and TCR loci) from this analysis. Chr13 was excluded from these analyses due to a pre-existing translocation on this chromosome.

SUPPLEMENTAL REFERENCE

Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nature genetics *43*, 1059-1065.

SUPPLEMENTARY TABLE 1: HTGTS LIBRARIES

26x I-SceI/ATM$^{-/-}$

Clone Name	Integration site	Primer	Treatment	Total hits	Hits on break chr		Hits on IgK		Hits on Ig/TCR loci	
					n	%	n	%	n	%
Chr2 $^{I\text{-}SceI}$	Chr2-180173673	5'	TA+STI	822	178	21.65	136	16.55	176	21.41
Chr2 $^{I\text{-}SceI}$	Chr2-180173673	5'	TA+STI+IR	9834	4407	44.81	94	0.96	126	1.28
Chr7 $^{I\text{-}SceI}$	Chr7-31203181	5'	TA+STI	467	61	13.06	67	14.35	82	17.56
Chr7 $^{I\text{-}SceI}$	Chr7-31203181	5'	TA+STI+IR	3005	802	26.69	37	1.23	47	1.56
Chr18 $^{I\text{-}SceI}$	Chr18-70688564	5'	TA+STI	1378	326	23.66	346	25.11	425	30.84
Chr18 $^{I\text{-}SceI}$	Chr18-70688564	5'	TA+STI+IR	8591	2862	33.31	279	3.25	331	3.85
Chr3 $^{I\text{-}SceI}$	Chr3-37646273	5'	TA+STI	27	16	59.26	2	7.41	2	7.41
Chr3 $^{I\text{-}SceI}$	Chr3-37646273	5'	TA+STI+IR	168	55	32.74	2	1.19	2	1.19
Chr2A $^{I\text{-}SceI}$	Chr2-152188613	5'	TA+STI	27	7	25.93	2	7.41	3	11.11
Chr2A $^{I\text{-}SceI}$	Chr2-152188613	5'	TA+STI+IR	148	44	29.73	0	0	1	0.68
Chr13 $^{I\text{-}SceI}$	Chr13-76395508	5'	TA+STI	32	23	71.88	2	6.25	4	12.50
Chr13 $^{I\text{-}SceI}$	Chr13-76395508	5'	TA+STI+IR	170	54	31.95	1	0.59	2	1.18
Total				24669	8835	35.81	968	3.92	1201	4.87

25x I-SceI/WT

Clone Name	Integration site	Primer	Treatment	Total hits	Hits on break site		Hits on IgK		Hits on Ig/TCR loci	
					n	%	n	%	n	%
Chr2B $^{I\text{-}SceI}$	Chr2-148388069	5'	TA+STI+ATMi	1839	164	8.92	93	5.06	111	6.04
Chr2B $^{I\text{-}SceI}$	Chr2-148388069	5'	TA+STI+ATMi+IR	1284	452	35.2	34	2.65	42	3.27
Chr2B $^{I\text{-}SceI}$	Chr15-5565929	5'	TA+STI	1815	117	6.45	2	0.11	4	0.22
Chr15 $^{I\text{-}SceI}$	Chr15-5565929	5'	TA+STI+ATMi	3575	179	5.01	85	2.38	104	2.91
Chr15 $^{I\text{-}SceI}$	Chr15-5565929	5'	TA+STI+IR	2489	243	9.76	0	0	5	0.2
Total				11002	1155	10.50	214	1.95	266	2.42

DEL-CJ/F1

Library Name	Integration site	Treatment	Total hits	Total SNP hits	Hits on break site chromosome		Hits on non-break site chromosomes	
					129	BALB/c	129	BALB/c
YZ198	Chr7-133846876	STI+ATMi+IR	3632	344	8	55	151	130
YZ212	Chr7-133846876	STI+ATMi+IR	12322	1184	18	76	446	565
YZ224	Chr7-133846876	STI+ATMi+IR	5448	561	14	68	218	261
YZ196	Chr9-75863761	STI+ATMi+IR	3227	301	86	12	89	114
YZ211	Chr9-75863761	STI+ATMi+IR	6160	469	126	10	149	184
YZ223	Chr9-75863761	STI+ATMi+IR	3919	380	106	18	111	145
Total			34708	3239	358	239	1164	1399

Supplementary Table 2: Hi-C libraries.

26x I-SceI/ATM$^{-/-}$

Hi-C Data Cell Line	Biological Replicate	Unique Interaction Pairs
Chr2 $^{I\text{-}SceI}$	1	45,122,598
Chr2 $^{I\text{-}SceI}$	2	49,316,157
Chr7 $^{I\text{-}SceI}$	1	34,679,715
Chr18 $^{I\text{-}SceI}$	1	16,553,041
Chr18 $^{I\text{-}SceI}$	2	38,046,959
Total		183,718,470

26x I-SceI/WT

Hi-C Data Cell Line	Biological Replicate	Unique Interaction Pairs
Chr2 $^{I\text{-}SceI}$	1	69,572,672
Chr15 $^{I\text{-}SceI}$	1	13,734,501
Total		83,307,173173