Extremely long range chromatin loops link topological domains to facilitate a diverse antibody repertoire

Lindsey Montefiori, Robert Wuerffel, Damian Roqueiro, Bryan Lajoie, Changying Guo, Tatiana Gerasimova, Supriyo De, William Wood, Kevin G. Becker, Job Dekker, Jie Liang, Ranjan Sen, and Amy L. Kenter

SUPPLEMENTAL FIGURES

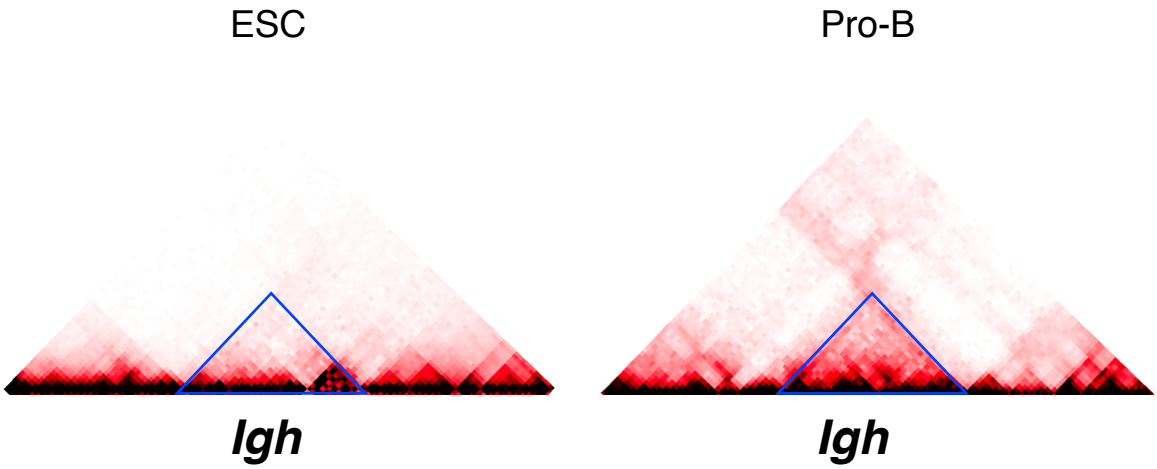

Figure S1. The *Igh* locus forms a 2.9 Mb compartment in an AMuLV pro-B cell line. (Related to Figure 1A). Normalized HiC interaction frequencies from an AMuLV pro-B cell line (Zhang et al., 2012) and ES cells (ESC) (Dixon et al., 2012) representing a segment of chromosome 12 (111,600,000-120,100,000; mm9) is displayed as a 2D heatmap binned at 100 kb resolution. HiC sequencing reads are indicated by the color (red) intensity from interacting pairs of Hind III fragments. The blue triangle spans the *Igh* locus. Compartments are defined as large chromosomal domains that preferentially interact with other compartments. In contrast, TADs are defined as self-associating and do not associate beyond their own boundaries.

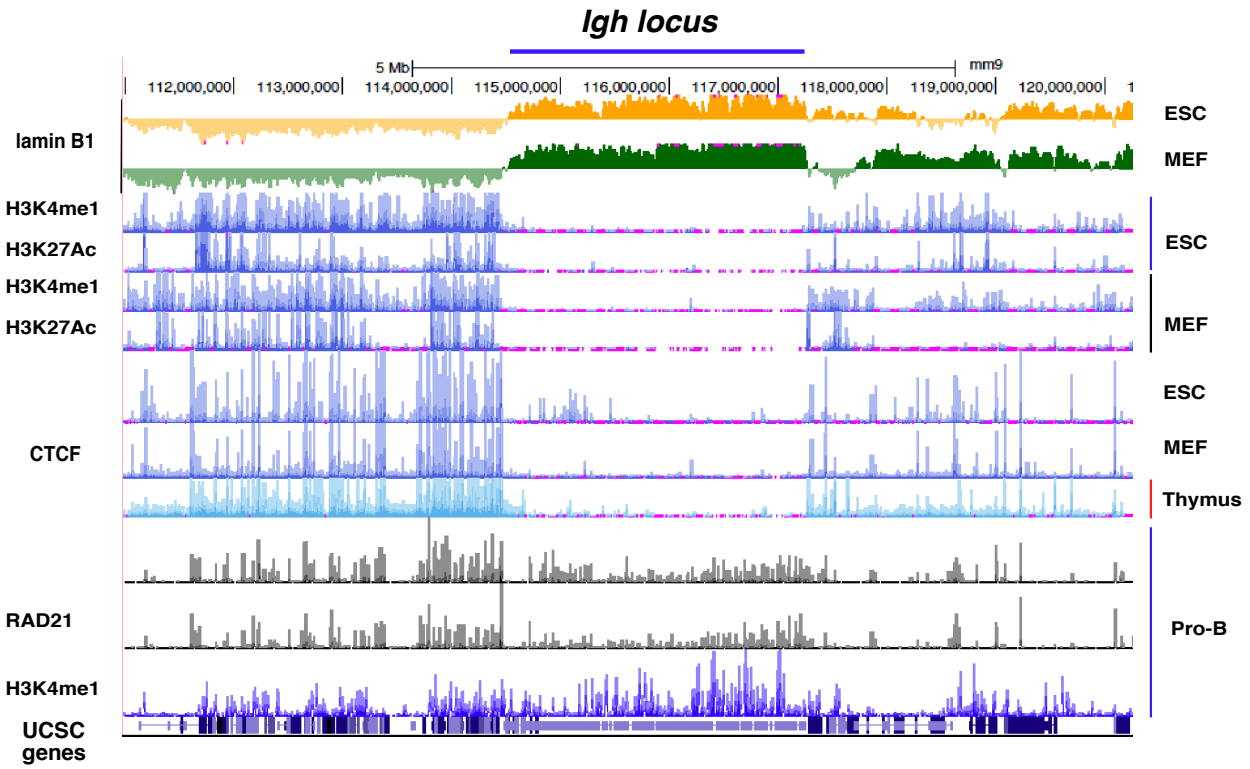

Figure S2. Differential chromatin modifications and CTCF binding at the Igh locus at pro-B- and non-B cells. (Related to Figure 1). Lamin associated domain (LAD) is detected the Igh locus in ES cells (ESC) and MEF but not in pro-B cells. Histone modifications, CTCF, Rad21 and laminB1 binding (DamID) of a segment of chr 12 including the Igh locus are shown. Genomic coordinates are as indicated. ChIP-seq and DamID data was taken from public data bases and summarized in Table S1.

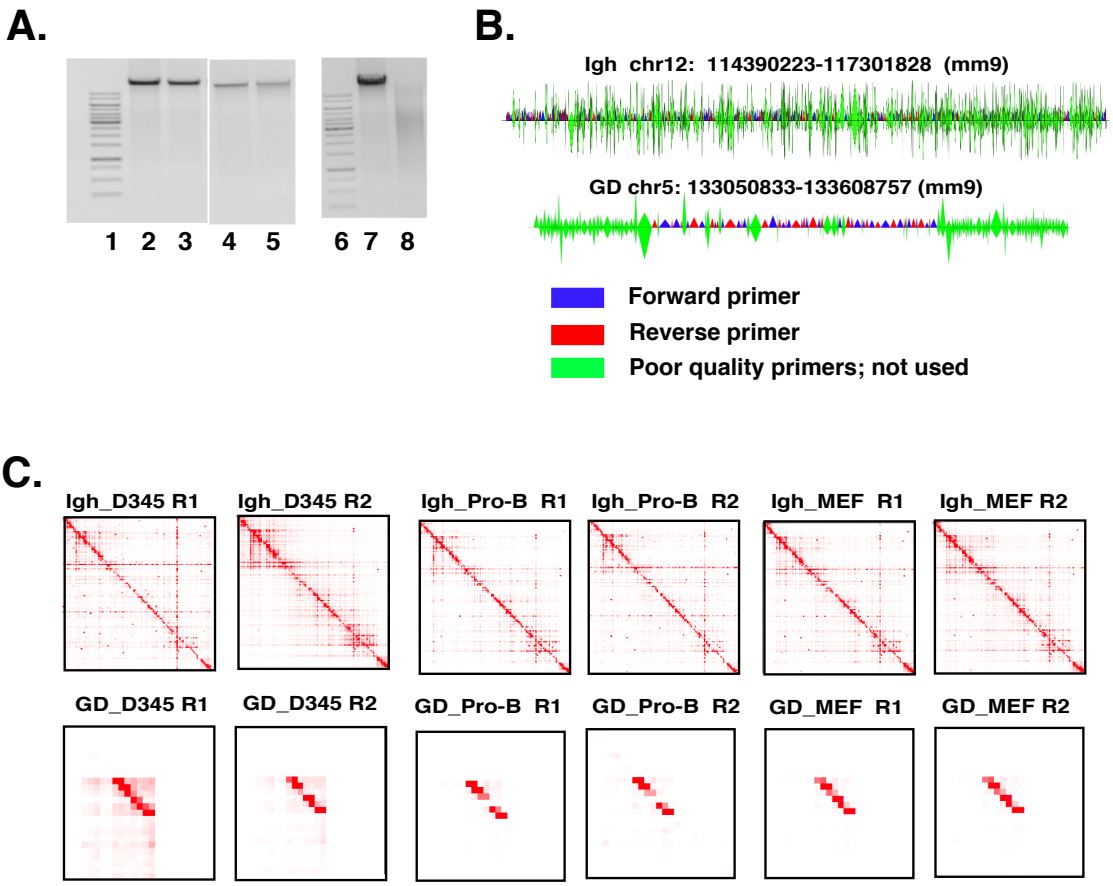

Figure S3. 5C library construction. (Related to Figure 1, 2) A) 3C templates (lanes 2-5, 7) and MW standards (lanes 1,6) were resolved on a 0.8% agarose gel. The bands at ~10kb indicate a relatively high restriction digest efficiency and ligation efficiency. 3C templates are from mature splenic B cells activated with LPS (lane 2), as described (Kumar et al., 2013) and is used as an internal control for quality, the D345 cell line (lane 3), and two replicates of Rag deficient pro-B cells (lanes 4,5). The presence of the 10 kb band is dependent on ligation (lane 7) and is absent when ligase is not added (lane 8). **B)** Alternating 5C primer design approach for the Igh locus and the gene desert on chr 5 and genomic coordinates are as indicated. The position of forward, reverse and poor quality primers are as shown. C) Raw read heatmaps for two replicates (R) of the D345 cell line, two replicates of Rag deficient pro-B cells, two replicates of MEFs are displayed for the Igh locus (top) and the chr 5 gene desert (bottom). Red intensity indicates read counts.

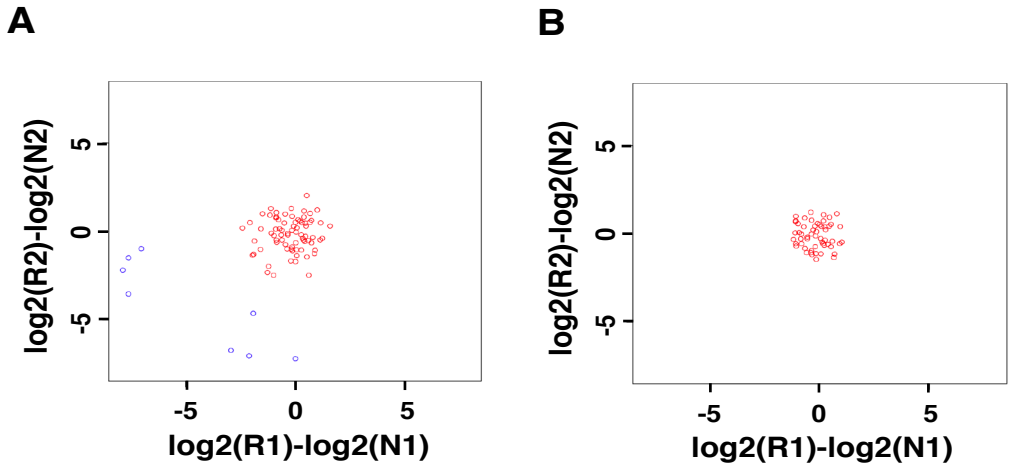

Figure S4. Identification of well correlated 5C primer pairs for the chr5 gene desert. (Related to Figure 2C,D). A normalizing coefficient from the chromosome 5 gene desert 5C data set is computed. The log2 ratio of reads for each primer pair is evaluated for biological replicates of MEF and pro-B cells and is used to identify and retain well behaved primers (see Suppl. Methods). The chr 5 primer pairs before correction (**A**) and after correction (**B**). Outliers are shown in blue and well behaved primers are shown in red.

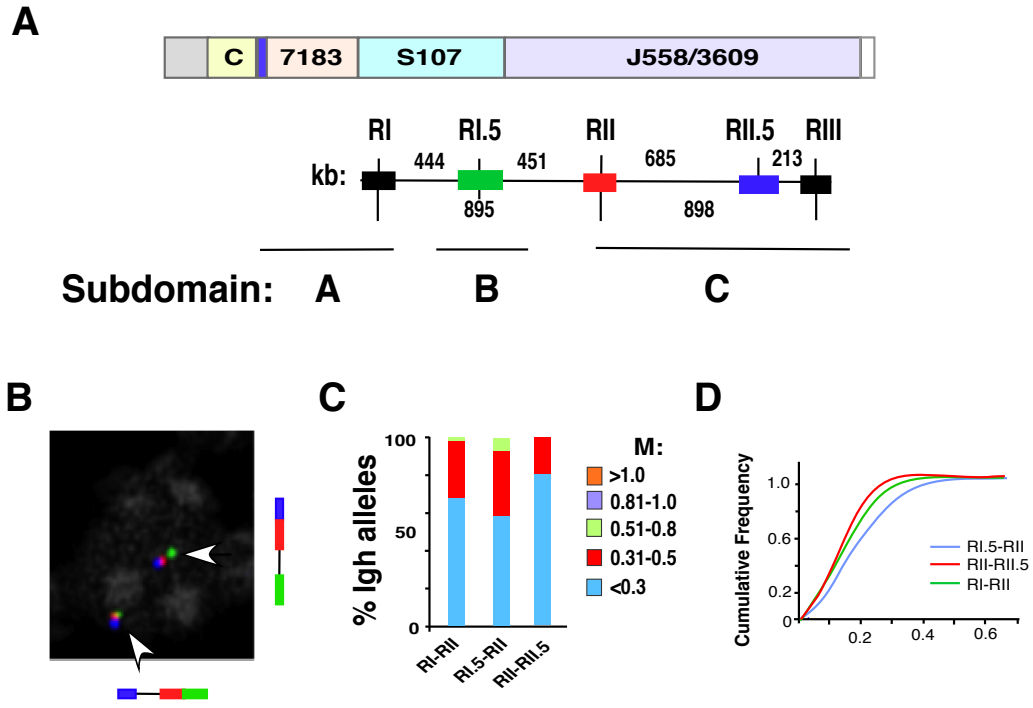

Figure S5. Sites I-III function as loop specific anchors. (Related to Figure 3). Statistical significance for probe distances is indicated in Table S6. **A)** A diagram of the Igh locus is shown indicating the location and distances between the three FISH probes that were labeled as follows: RI.5 (green), RII (red) and RIII (blue). The FISH probes RI and RIII are indicated in black and were used in parallel studies. The topological sub-domains A, B and C are arrayed at bottom and were identified in 5C studies shown in Figure 2. **B)** A representative three color FISH using RI.5 (green) RII (green) and RIII (blue) and purified Rag deficient pro-B cells is shown. Purified Rag deficient pro-B cells from at least three mice were used. Data are from one independent experiment. **C)** Quantitation of FISH data. The distance between the red, green and blue probes shown in (B) for ~100 nuclei were divided into 5 categories: <0.3, 0.31-0.5, 0.51-0.8, 0.81-1.0 mm. Inter-probe distances for RI-RII were taken from the studies presented in Figure 3C. The percentage of Igh alleles in each category was determined (y axis) for each probe combination and is represented in different colors. Probe combinations tested are indicated along the x-axis. Probe combinations are shown below the histograms. **D)** Cumulative frequency analyses are shown for each probe combination.

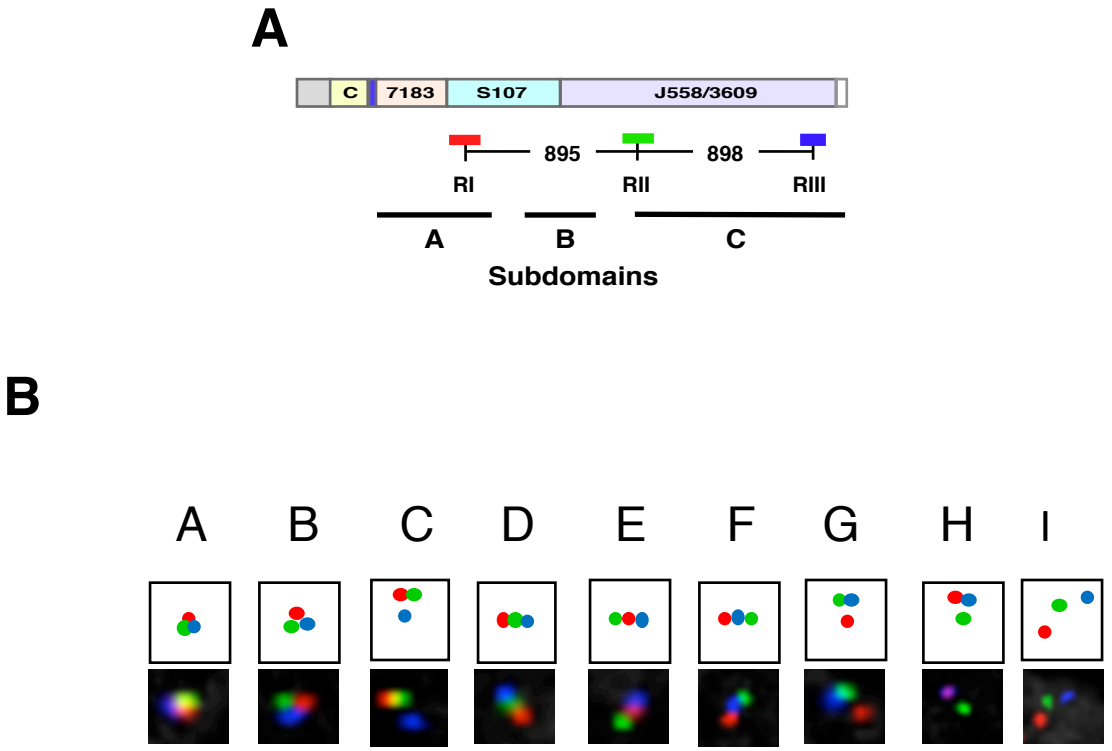

Figure S6. Guidelines used to score allele configurations in three color FISH assays (related to Figure 4). A) A schematic of the WT *Igh* locus with genomic distances between BAC probes. The probes combinations are RI (red), RII (green) and RIII (blue). The topological sub-domains A, B and C that are derived from the 5C studies shown in Figure 2 are arrayed at bottom. **B)** BAC probes RI, RII and RIII, labeled as indicated, were hybridized simultaneously to Rag2 deficient pro-B cells. The percentage of *Igh* alleles in various spatial configurations was determined as follows: inset A) All pair-wise distances less than 0.25 µm; inset B) All pair-wise distances 0.25-0.5 µm, insets C,G,H) Paired probes < 0.3 µm; separated probe > 0.3 µm from both paired probes. Insets D-F) Linear configuration, < 0.3 µm between neighboring probes. Inset I) All pair-wise distances > 0.5 µm.

Supplementary Tables

Table S1. Source of ChIP-seq data from MEF and ESC[1]

Source	Description	Category	ChIP	Build	Source lab	Documents	Data Source
MEF	ChIP-seq	Primary cells	H3K27Ac	mm9	Ren	/gbdb/mm9/bbi/wgEncodeLicrHistoneMefH3k27acMAdult8wksC57bl6StdPk	http://genome.ucsc.edu
MEF	"	"	H3K4me1	"	"	/gbdb/mm9/bbi/wgEncodeLicrHistoneMefH3k4me1MAdult8wksC57bl6StdPk	"
MEF	"	"	CTCF	"	"	/gbdb/mm9/bbi/wgEncodeLicrTfbsMefCtcfMAdult8wksC57bl6StdSig.bigWig	"
MEF	DamID	"	Lamin B1	"	van Steensel	GEO accession number GSE17051	ncbi.nlm.nih.gov/geo/
ES-Bruce4	ChIP-seq	Cell line	H3K27Ac	"	Ren	/gbdb/mm9/bbi/wgEncodeLicrHistoneEsb4H3k27acME0C57bl6StdSig.bigWig	http://genome.ucsc.edu
ES-Bruce4	"	"	H3K4me1	"	"	/gbdb/mm9/bbi/wgEncodeLicrHistoneEsb4H3k4me1ME0C57bl6StdSig.bigWig	"
ES-Bruce4	"	"	CTCF	"	"	/gbdb/mm9/bbi/wgEncodeLicrTfbsEsb4CtcfME0C57bl6StdSig.bigWig	"
Embryonic stem cells	DamID	Primary cells	Lamin B1	"	van Steensel	GEO accession number GSE17051	ncbi.nlm.nih.gov/geo/
Thymus	ChIP-seq	"	CTCF	"	Ren	/gbdb/mm9/bbi/wgEncodeLicrTfbsThymusCtcfMAdult8wksC57bl6StdSig.bigWig	http://genome.ucsc.edu
BM Pro-B	"	"	H3K4me1	"	Murre	GEO accession number GSM537991	
"	"	"	CTCF	"	"	GEO accession number GSE40173	ncbi.nlm.nih.gov/geo/
"	"	"	RAD21	"	"	GEO accession number GSE40173	"
							"

1 Related to Fig. 1A and Fig. S1

Table S2. 5C library statistics[1,2]

MatrixFile	Library Name	Total Reads	Total Interactions	Total Cis	Total Trans
4d892a926efeea0f58667a1e465d2147.5C.interactions	AKIghExGD-B6-R1	35,830,368	15,853	12,798	3,055
30737ec08f6743274280abf13f4c661c.5C.interactions	AKIgHExGD-Cy-R1	9,325,281	17,940	12,783	5,157
2d97cea72765fa51aac4a5a1184d92e1.5C.interactions	AKIgHExGD-B-ProB-R1*	22,049,248	15,357	12,556	2,801
5e4f7a42d4c93aa4fe2df836462dd7e7.5C.interactions	AKIgHExGD-B-ProB-R2*	22,300,921	14,508	11,341	3,167
f62322f874bf1ed51dd950b2dc3ee13d.5C.interactions	AKIgHExGD-B-Mef-R1	24,191,175	15,488	12,490	2,998
1dfaad096b058512adf4c67387e66665.5C.interactions	AKIgHExGD-B-Mef-R2	22,868,398	15,426	12,536	2,890

1 Some data from these files was previously reported (Kumar et al., 2013).

2 Related to Fig. 1B, 2

Table S3. Summary of 5C paired-end reads aligned using Bowtie[1]

Cell	Replicate	Number of reads						
		Original	Mapped PE1	%	Mapped PE2	%	Matched	%
pro-B	1	25,175,204	23,034,072	92%	22,593,323	90%	21,424,015	85%
	2	25,919,640	23,420,680	90%	23,024,139	89%	21,964,806	85%
MEF	1	26,537,062	24,939,514	94%	24,651,205	93%	23,858,581	90%
	2	24,766,329	23,358,067	94%	23,151,825	94%	22,332,216	90%

1 Related to Fig. 2 C,D

Table S4. Correlation analysis of 5C reads for the chr 5 gene desert

Cell	Replicate	Cell	Replicate	coefficient	p-value
pro-B	1	MEF	1	0.781	1.10e-19
	2		1	0.772	5.02e-18
	1		2	0.923	2.50e-38
	2		2	0.880	1.42e-28

1 Related to Fig. 2 C,D

Table S5. Scaling coefficients obtained from all combinations of 5C library replicates[1]

Set	Cell/replicate	Cell/replicate	Coefficient
1	MEF1	pro-B1	1.388893
	MEF2	pro-B2	1.784458
2	MEF1	pro-B2	1.367990
	MEF2	pro-B1	1.801393
3	MEF1	pro-B1	1.388893
	MEF1	pro-B2	1.402300
4	MEF2	pro-B1	1.801015
	MEF2	pro-B2	1.784458
5	MEF1	pro-B1	1.352814
	MEF2	pro-B1	1.961408
6	MEF1	pro-B2	1.364803
	MEF2	pro-B2	1.834602

1 Related to Fig. 2 C,D

Table S6. FISH Probe Associations are Statistically Significant[1]

Figure	Experiment	p-value[2] (K-S test)	p-value[2] (Wilcoxon test)	Alleles measured [3] (first probe-pair, second probe-pair)
3C,D	RI-RII vs H14-RI (pro-B)	9.57E-06	4.45E-08	100,300
3F,G	RI-RII (pro-B vs non-B)	< 2.2E-16	< 2.2E-16	300,100
	RII-RIII (pro-B vs non-B)	1.85E-09	1.64E-10	200,100
	RI-RIII (pro-B vs non-B)	< 2.2E-16	< 2.2E-16	200,100
5C,D	RI-RII (P-E- vs P-E+)	0.9996	0.5705	100,100
	RII-RIII (P-E- vs P-E+)	0.281	0.2059	100,100
	RI-RIII (P-E- vs P-E+)	0.2106	0.1462	100,100
5F,G	RI-RII (Pax5+/+ vs Pax5-/-)	0.01094	0.04447	120,120
	RII-RIII (Pax5+/+ vs Pax5-/-)	1.79E-09	3.24E-12	120,120
	RI-RIII (Pax5+/+ vs Pax5-/-)	4.09E-09	8.34E-10	120,120
S4C	RI-RII vs RI.5-RII (pro-B)	0.09176	0.01161	100,100
	RI-RII vs RII-RII.5 (pro-B)	0.2303	0.2102	100,100
	RI.5-RII vs RII-RII.5 (pro-B)	0.006302	0.001674	100,100

1 Related to Figures 3,4,5 and Figure S5.
2 P-values are reported for each experiment comparing the spatial distance between two probes. For each comparison, two tests for significance were performed, the Kolmogorov-Smirnov test (K-S) and the Wilcoxon test.
3 The number of alleles measured refers respectively to the two sets of probe pairs as they appear in the Experiment column.

Supplemental Experimental Methods

Statistical analyses

Statistical analyses were done using a two-tailed Student's t-test except where indicated. FISH studies were analyzed for statistical significance using the Wilcoxon and Kolmogorov-Smirnov tests (Table S6) (See also Figure 3,4,5).

5C primer design

5C primers follow an alternating pattern and were designed at HindIII restriction sites using the 5C primer design tools (Dostie et al., 2006) that are available online (http://my5C.umassmed.edu) (Lajoie et al., 2009) were previously reported (Kumar et al., 2013). 5C primers located at highly repetitive sequences were excluded (Fig. S3B). For the *Igh* locus, a total of 112 forward and 113 reverse primers were designed for all HindIII restriction fragments between coordinates chr12:114390223-117301828 (mm9), a 2.91 Mb region, and produce a total of 12,656 *cis* interactions. A chromosome 5 gene desert (chr5:133194370-133464774 (mm9)) spanning 270 kb lacks expressed genes, ORF mRNAs and microRNAs was analyzed by 5C using 24 forward and 25 reverse 5C primers for a total of 525 potential cis interactions. Primers settings were: U-BLAST: 3; S-BLAST: 80: 15-MER: 800; MIN_FSIZE: 1000; MAX_FSIZE: 15,000; OPT_TM: 65; OPT_PSIZE: 30. Several primers with somewhat relaxed parameters were included to provide a more complete coverage of the *Igh* locus. The universal T7 sequence (5'-TAATACGACTCACTATAGCC-3') was tethered to all forward primers. The reverse complement to the universal T3 sequence (5'TATTAACCCTCAACTAAAGGGA-3') was tethered to all reverse primers. Each reverse primer was modified with a phosphate group at the 5' end using T4 polynucleotide kinase.

5C libraries construction and Paired-end Solexa Sequencing

5C library construction was performed as described (Dostie and Dekker, 2007; Kumar et al., 2013) using 3C chromatin libraries. 3C chromatin was prepared from $Rag2^{-/-}$ pro-B cells that were purified from the bone marrow by positive selection using anti-CD19-coupled magnetic beads (Stem Cell Technologies, BC, Canada) and then expanded on OP9 cells as described (Pongubala et al., 2008). In 3C based assays, differences in formaldehyde crosslinking efficiency, restriction enzyme digestion, ligation efficiency, PCR amplification, Illumina adaptor ligation and sequencing can complicate sample to sample comparison of 5C data sets. To control for these factors, the frequency of 5C reads associated with the chromosome 5 gene desert were used to derive a scaling factor that is then used to normalize 5C *Igh* data sets from MEF and pro-B cells. The integrity of the 3C chromatin libraries used for 5C libraries were confirmed (Fig. S3A). Primers were pooled so that each primer was present at a equimolar concentration of 0.75 fmol/ml and annealed with 700 ng of 3C template in a final volume of 10 ml. Negative controls demonstrating the specificity of the 5C library included carrying out the reaction in the absence of 5C primers, ligase, or 3C template. 5C products were analyzed by Solexa paired-end deep sequencing. A virtual genome of all potential 5C products was created using DNA sequence adjacent to Hind III restriction fragment sites. Reads were aligned to the virtual genome using Novoalign (http://www.novocraft.com) (also see Fig. 1B, 2A). Each paired-end read was mapped independently and only paired-ends for which one end mapped to a forward primer and the other to a reverse primer were considered. Reads that mapped to more than one location were discarded. The 5C library composition for each experiment shown in Figures 1B and 2 is summarized in Tables S2,S3. 5C analyses were carried out on two biological replicates from MEF cells or Rag2-/- pro-B cells. We obtained 22,049,248 and 22,300,921 mapped reads for two biological replicates of Rag2-/- pro-B cells of which 21,927,201 and 21,689,397 could be specifically mapped back to interactions within the *Igh* locus using Novoalign (http://www.novocraft.com), respectively, as previously reported (Kumar et al., 2013). For MEF cells we obtained 24191175 and 22868398 mapped reads for two biological replicates of which 23487611 and 22405605 could be specifically mapped back to the *Igh* locus. In all instances the mappable reads were proportional to the degree of multiplexing, indicating equivalent library quality despite different read numbers. The heat maps are scaled as follows: for Figure 1B, MEF 37-1840, Rag2-/- pro B cells: 19-1254, D345: 89-6355 and for Figure 2A, right panel, Rag2-/- pro B-MEF: 19-1254. Raw 5C data and a complete 3C protocol are available by request.

Direct sample-to-sample comparison of 5C data sets

To perform direct sample to sample comparison of 5C data sets the frequency of 5C reads associated with the chromosome 5 gene desert were used to derive a scaling factor that is then used to normalize 5C *Igh* data sets from MEF and pro-B cells. The chr5 gene desert was chosen for this purpose since it is devoid of expressed genes, mRNAs and fails to express open reading frame mRNAs and therefore is unlikely to alter its chromatin state at different stages of

development. Only primers separated by a distance of at least 12 kb, to exclude random proximity events, and <100 Kb are considered as previously described (Dekker, 2006). A virtual genome of all potential 5C reads was created and aligned using Bowtie (Langmead et al., 2009). The 5C library composition for each experiment shown in Figure 1B and 2B,C is summarized in Table S2. Each paired-end read was mapped independently and only paired-ends for which one end mapped to a forward primer and the other to a reverse primer were considered. The majority of primer pairs produced read frequencies that were well correlated between biological replicates and between samples (Table S4). However a minority of primer pairs for the chromosome 5 gene desert were not well correlated and these outliers were removed as described below (Fig. S4, marked in blue). We used an iterative process to eliminate gene desert primers that do not agree well between biological replicates. 1) For every possible combination of replicates k_1 and k_2, we compute log2-ratios of pro-B over MEF. 2) Primer pairs that vary more than 2 standard deviations from the mean are excluded. 3) The remaining primers are then used to compute the normalizing coefficients and to correct the interaction counts of the gene desert primers. 4) This process is repeated until all discordant primers have been eliminated. The four steps described above yield a subset of gene desert primers that are well correlated between replicates and cell types. The curated gene desert primers are then used to derive normalizing coefficients for the *Igh* 5C data sets.

Computing a normalizing coefficient from the chr5 gene desert

Here we compute the normalizing coefficients from the gene desert that enable comparison of 5C interaction frequencies at the *Igh* locus of MEF and pro-B cell samples. For a pair p of primers (i, j) in the chr5 gene desert we will have two interaction counts: $\boldsymbol{B}_{kp}$ and $\boldsymbol{M}_{kp}$, where $\boldsymbol{B}$ and $\boldsymbol{M}$ represent the matrices with interactions identified in pro-B and MEF cells respectively; p is the primer pair; and k is the specific combination of biological replicates. In our case, possible combinations are $k = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$ where the first element of each pair corresponds to pro-B and the second to MEF. This can be generalized to any problem with two cell types with R_1 replicates in cell 1 and R_2 replicates in cell 2. A normalizing factor for each replicate combination k can be computed as indicated in equation 1 [4].

$$n_k = \frac{\sum_{p=1}^{P} B_{kp}}{\sum_{p=1}^{P} M_{kp}} \qquad (1)$$

We want to normalize all the ratios $r_{kp} = \frac{B_{kp}}{M_{kp}}$ for all pairs of primers $p = \{1, 2, \ldots, P_{\text{desert}}\}$ and to obtain

corrected ratios $$\hat{r}_{kp} = \frac{B_{kp}}{M_{kp}} \cdot \frac{1}{n_k}$$

Ideally, the corrected ratio $\hat{r}_{1p}$, for example in replicate 1 of both cells, is expected to be very similar to the ratio $\hat{r}_{2p}$ in replicate 2, both for pair p [4, 5]. For all primer pairs p and replicates (1, 1) and (2, 2)

$$\hat{r}_{1p} = \frac{B_{1p}}{M_{1p}} \cdot \frac{1}{n_1} \qquad (2)$$

$$\hat{r}_{2p} = \frac{B_{2p}}{M_{2p}} \cdot \frac{1}{n_2} \qquad (3)$$

we expect

$$\frac{\hat{r}_{1p}}{\hat{r}_{2p}} = \frac{B_{1p}}{n_1 M_{1p}} \cdot \frac{n_2 M_{2p}}{B_{2p}} \approx 1$$

which is equivalent to

$$\log_2(\frac{\hat{r}_{1p}}{\hat{r}_{2p}}) \approx 0$$

Following the previous example, at the end of six iterative rounds of this process for replicates (1, 1) and (2, 2) the pairs of primers that show consistency for the log2 ratio of $\hat{r}_{1p}$ over $\hat{r}_{2p}$ are retained (Fig. S4, marked in red). The log2 ratio is used to identify and eliminate poorly behaved primers. To assess the efficiency of this process, Spearman correlation coefficients were calculated for all possible combinations of replicates between pro-B and MEF after primer correction. All replicates showed an excellent correlation between the interactions in the gene desert (Spearman correlation coefficient > 0.772 and p-value < 5.02e-18) (Table S4).

Calculating a scaling coefficient for comparison of 5C data sets

When computing $\log_2(\frac{\hat{r}_{1p}}{\hat{r}_{2p}})$ the value of $\hat{r}_{1p}$ is obtained from replicate 1 of both pro-B and MEF. In general, we can compute a corrected ratio $\hat{r}_{k_1 p}$ for any combination k_1 of replicates. Therefore, the iterative process that removes discordant primers in the gene desert must explore all possible $\log_2(\frac{\hat{r}_{k_1 p}}{\hat{r}_{k_2 p}})$ for sets of combinations k_1 and k_2. Table S5 shows the 6 possible sets of combinations. In this way, the final scaling coefficient for each replicate is computed as an average of all the coefficients computed with that replicate.

Scaling coefficient for pro-B, replicate 1 is

$$n_1 = 1.615736 = \frac{1.388893 + 1.801393 + 1.388893 + 1.801015 + 1.352814 + 1.961408}{6}$$

Scaling coefficient for pro-B, replicate 2 is

$$n_2 = 1.589768 = \frac{1.784458 + 1.367990 + 1.402300 + 1.784458 + 1.364803 + 1.834602}{6}$$

Thus, the total number of interactions in the gene desert of MEF is approximately 1.6 fold higher as compared to pro-B cells. We normalize the 5C data sets accordingly to enable sample to sample comparisons.

Normalizing the 5C *Igh* data sets

To compare a given primer pair *p* in pro-B versus MEF we compute the log2 fold-change of the interaction counts between cells types and use the average log2 ratios from both replicates:

$$v_p = \frac{\log_2(\hat{r}_{1p}) + \log_2(\hat{r}_{2p})}{2}$$

$$= \frac{1}{2}\log_2(\hat{r}_{1p}\,\hat{r}_{2p})$$

$$= \log_2(\sqrt{\hat{r}_{1p}\,\hat{r}_{2p}})$$

Using the normalization coefficients obtained in the previous section and substituting from equations (2) and (3)

$$v_p = \log_2\left(\sqrt{\frac{B_{1p}}{n_1 M_{1p}} \cdot \frac{B_{2p}}{n_2 M_{2p}}}\right)$$

$$= \log_2\left(\sqrt{\frac{B_{1p}B_{2p}}{M_{1p}M_{2p}} \cdot \frac{1}{n_1 n_2}}\right)$$

we obtain a measure of the log2 fold-change of primer pair *p* between the combined replicates pro-B and MEF

$$v_p = \log_2\left(\sqrt{\frac{B_{1p}B_{2p}}{M_{1p}M_{2p}}}\right) - \log_2(\sqrt{n_1 n_2}) \qquad (4)$$

The value v_p provides an objective measure of the difference in the number of interactions for primer pair *p* between the cells types. The value v_p is obtained by computing the geometric mean of replicates in pro-B $\sqrt{B_{1p}B_{2p}}$ and MEF $\sqrt{M_{1p}M_{2p}}$ and then correcting by the normalizing coefficients n_1 and n_2 computed for both replicates of pro-B derived above.

Equation (4) generalizes to an experiment with *m* replicates where the log2 fold-change for primer pair *p* is

$$v_p = \log_2\left(\sqrt{\frac{B_{1p}B_{2p}}{M_{1p}M_{2p}}}\right) - \log_2(\sqrt{n_1 n_2}) \qquad (5)$$

In order to avoid singularities in the operations, the following considerations must be made:

If for replicate i, $B_{ip} = 0$ then set $B_{ip} = 0.01$. The same applies to M_{ip}. If for primer p, $B_{ip} = 0$ and $M_{ip} = 0$ for all i, then set $v_p = 0$.

For each primer pair p, the value v_p with the log2 fold-change of interaction frequency between pro-B and MEF was used to classify the interactions as: 1) pro-B prevalent, 2) MEF prevalent or 3) constitutive (also see Figure 2C). A threshold of $v_p > 1.6$ is used to classify primer pairs in this category. The log2 value of 1.6 is equivalent to a 3-fold difference in normalized interactions For constitutive interactions the threshold is $1.6 \leq v_p \leq -1.6$.

Preparation of chromatin for 5C analyses by formaldehyde crosslinking will lead to some level of nonspecific crosslinking interactions. We used LOESS with a second-degree polynomial ($d = 2$) and a span a = 0.1 to assess the random background interactions. The parameter a is used to determine the fraction of points, or span, in the vicinity of the estimating point that will be used to compute the weight. A LOESS curve was computed for each normalized replicate and provided a point estimation of the expected number of interactions for different distances between primers. For each primer pair p we use the distance between the primers to subtract the expected LOESS count from the observed count. If the number of expected counts is larger than the observed counts, then we set the corrected interactions to zero.

Fluorescent In Situ Hybridization

Pax5-deficient pro-B cells (kindly provided by Dr. J. Pongubala) were expanded on S17 stromal cell cultures for FISH analyses (Pongubala et al., 2008). DNA probes for fluorescent in situ hybridization (FISH) were prepared from locus specific BACs (Invitrogen, CA). All genomic coordinates are chr12 (mm9). Probes were RI (115051557-115227487) (BAC 373N4), RI.5 (115,549,000-115,629,000) (BAC 365C11), RII (115944024-116124641) (BAC 70F21), RII.5 (116553036-116777349) (BAC 451H16), RIII (116777388-117011222) (BAC 368C22 or 230L2). BAC RP23-201H14 was kindly provided by Dr. C. Murre (UCSD). Nick translation was used to label BAC probes in the presence of Alexa Fluor 594 (red) and 488 (green) and BAC RP23-201H14 was labeled with Alexa Fluor 697 (blue). Labeled probes were hybridized with fixed cells and FISH was performed as described (Guo et al., 2011). Serial optical sections (n=30-40; each slice is 0.1 mm thick) spaced by 0.2 mm were acquired. NIS-Elements software (Nikon, NY) was used to deconvolute data sets and optical sections were merged to produced 3D images. Spatial distances between probes were measured as previously described (Jhunjhunwala et al., 2008). P values of statistical significance were calculated using Wilcoxon and Kolmogorov-Smirnov tests (Table S6).

SUPPLEMENTAL REFERENCES

Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. Nat Methods *3*, 17-21.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Dostie, J., and Dekker, J. (2007). Mapping networks of physical interactions between genomic elements using 5C technology. Nat Protoc *2*, 988-1002.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C.*, et al.* (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res *16*, 1299-1309.

Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R., Oltz, E.M., and Sen, R. (2011). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. Cell *147*, 332-343.

Jhunjhunwala, S., van Zelm, M.C., Peak, M.M., Cutchin, S., Riblet, R., van Dongen, J.J., Grosveld, F.G., Knoch, T.A., and Murre, C. (2008). The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. Cell *133*, 265-279.

Kumar, S., Wuerffel, R., Achour, I., Lajoie, B., Sen, R., Dekker, J., Feeney, A.J., and Kenter, A.L. (2013). Flexible ordering of antibody class switch and V(D)J joining during B-cell ontogeny. Genes Dev *27*, 2439-2444.

Lajoie, B.R., van Berkum, N.L., Sanyal, A., and Dekker, J. (2009). My5C: web tools for chromosome conformation capture studies. Nat Methods *6*, 690-691.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Pongubala, J.M., Northrup, D.L., Lancki, D.W., Medina, K.L., Treiber, T., Bertolino, E., Thomas, M., Grosschedl, R., Allman, D., and Singh, H. (2008). Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5. Nat Immunol *9*, 203-215.

Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell *148*, 908-921.