Immunity, Volume *39*

Supplemental Information

Flexible Long-Range Loops in the V_H Gene Region of the *Igh* Locus Facilitate the Generation of a Diverse Antibody Repertoire

Jasna Medvedovic, Anja Ebert, Hiromi Tagoh, Ido M. Tamir, Tanja Schwickert, Maria Novatchkova, Qiong Sun, Pim J. Huis in 't Veld, Chunguang Guo, Hye Suk Yoon, Yves Denizot, Sjoerd J.B. Holwerda, Wouter de Laat, Michel Cogné, Yang Shi, Frederick W. Alt, and Meinrad Busslinger

Supplemental Inventory

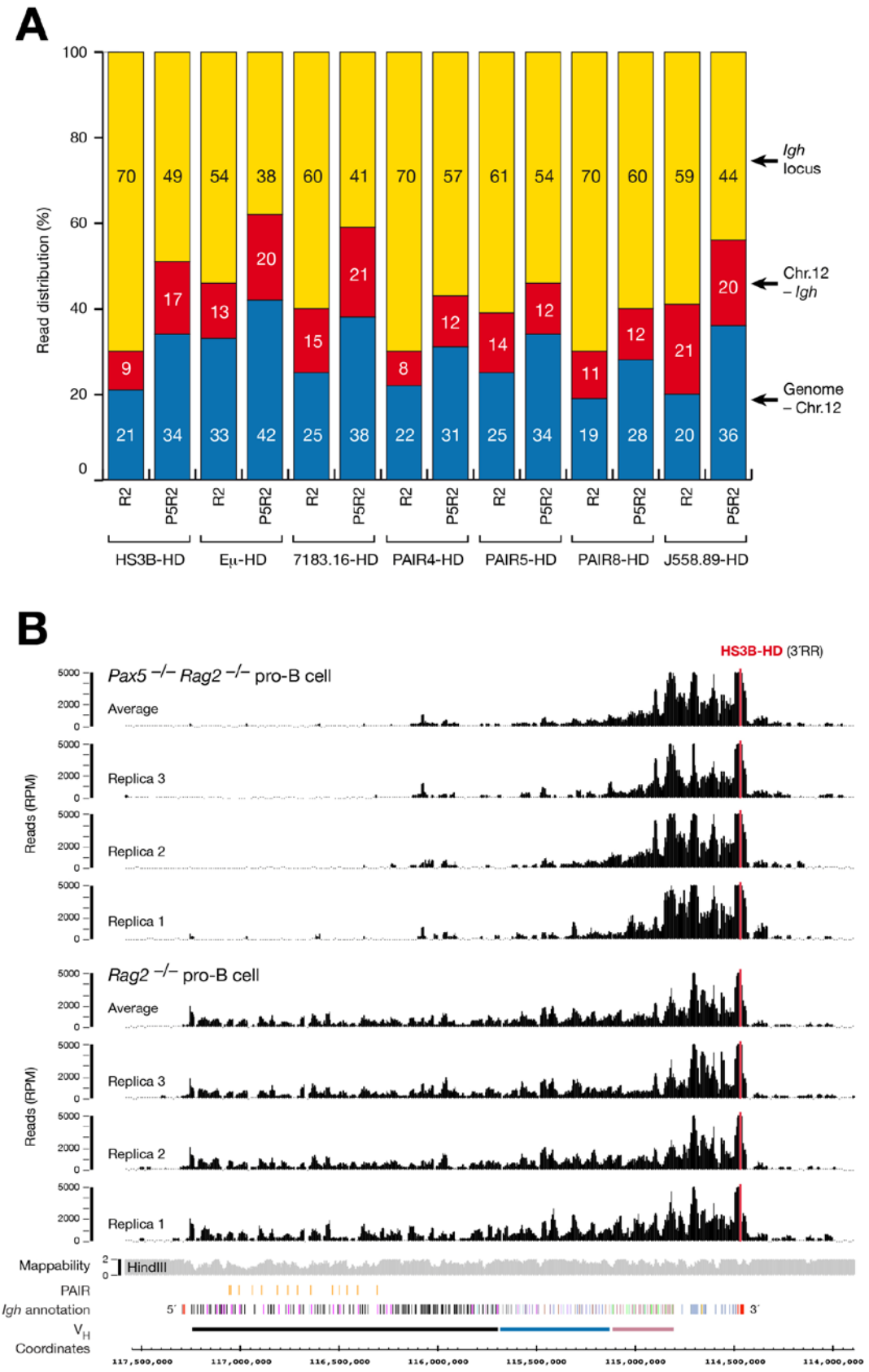
A
Read distribution (%)
100
80
60
40
20
0
70 49 54 38 60 41 70 57 61 54 70 60 59 44
9 17 13 20 15 21 8 12 14 12 11 12 21 20
21 34 33 42 25 38 22 31 25 34 19 28 20 36
Igh locus
Chr.12 – Igh
Genome – Chr.12
R2 P5R2 R2 P5R2 R2 P5R2 R2 P5R2 R2 P5R2 R2 P5R2 R2 P5R2
HS3B-HD
Eμ-HD
7183.16-HD
PAIR4-HD
PAIR5-HD
PAIR8-HD
J558.89-HD
B
HS3B-HD (3′RR)
Pax5 −/− Rag2 −/− pro-B cell
Average
Replica 3
Replica 2
Replica 1
Rag2 −/− pro-B cell
Reads (RPM)
5000
2000
0
Mappability
HindIII
PAIR
Igh annotation
5′
3′
V_H
Coordinates
117,500,000
117,000,000
116,500,000
116,000,000
115,500,000
115,000,000
114,500,000
114,000,000

Figure S1

Figure S1. Genomic read distribution and reproducibility of 4C-seq experiments.
(A) Genomic distribution of the 4C-seq reads obtained with the indicated viewpoints. The reads mapping to the *Igh* locus, the rest of chromosome 12 (Chr.12 – *Igh*) or the entire genome minus chromosome 12 (Genome – Chr.12) are shown as percentages of all mapped sequence reads. The average values of 3 independent 4C-seq experiments are indicated for each viewpoint and pro-B cell type of the *Rag2*$^{-/-}$ (R2) or *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ (P5R2) genotype. (B) Reproducibility of the 4C-seq results. Three independent 4C-seq experiments were performed with short-term cultured *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells, and the individual and average interaction patterns of the HS3B-HD viewpoint are shown for both cell types.

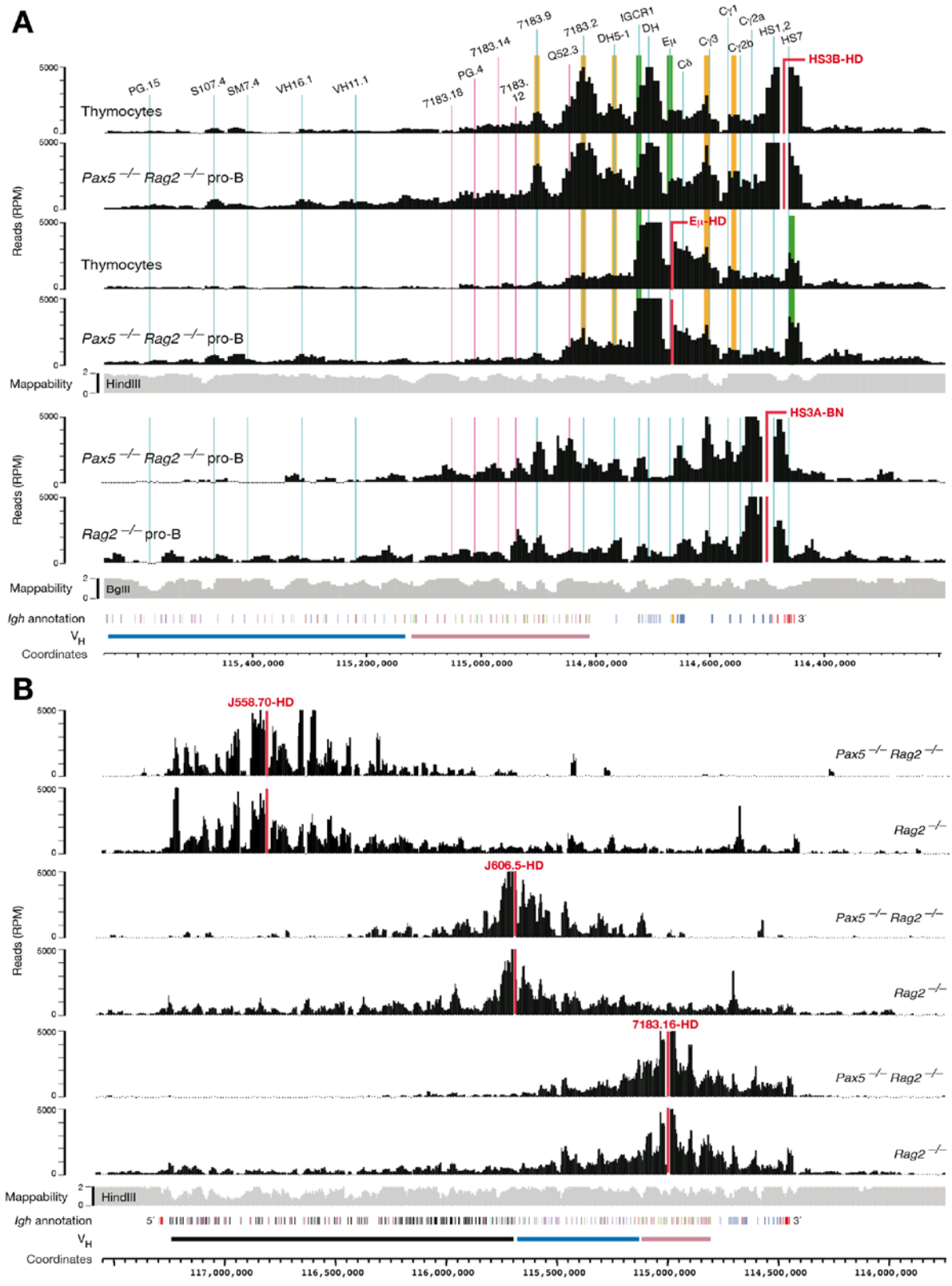
A
PG.15
S107.4
SM7.4
VH16.1
VH11.1
7183.18
PG.4
7183.14
7183.12
7183.9
Q52.3
7183.2
DH5-1
IGCR1
DH
Eμ
Cδ
Cγ3
Cγ1
Cγ2b
Cγ2a
HS1,2
HS7
HS3B-HD
Thymocytes
Pax5 −/− Rag2 −/− pro-B
Eμ-HD
Thymocytes
Pax5 −/− Rag2 −/− pro-B
Reads (RPM)
Mappability
HindIII
HS3A-BN
Pax5 −/− Rag2 −/− pro-B
Rag2 −/− pro-B
Reads (RPM)
Mappability
BglII
Igh annotation
5′
3′
VH
Coordinates
115,400,000
115,200,000
115,000,000
114,800,000
114,600,000
114,400,000
B
J558.70-HD
Pax5 −/− Rag2 −/−
Rag2 −/−
J606.5-HD
Pax5 −/− Rag2 −/−
Rag2 −/−
7183.16-HD
Pax5 −/− Rag2 −/−
Rag2 −/−
Reads (RPM)
Mappability
HindIII
Igh annotation
5′
3′
VH
Coordinates
117,000,000
116,500,000
116,000,000
115,500,000
115,000,000
114,500,000
114,000,000

Figure S2

Figure S2. 4C-seq interaction patterns of selected viewpoint across the *Igh* locus.
(A) Interaction patterns of the proximal viewpoints Eμ-HD, HS3A-BN and HS3B-HD (red) with sequences of the *Igh* 3' region in ex vivo isolated thymocytes and short-term cultured *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells. Vertical lines indicate the positions of V_H, D_H and C_H gene segments as well as the regulatory IGCR1, Eμ, 3'RR (HS1,2) and 3'CBE (HS7) regions. Previously known and newly identified interactions are highlighted in green and orange, respectively. (B) Long-range interactions of the viewpoints J558.70-HD, J606.5-HD and 7183.16-HD (red) along the *Igh* locus in short-term cultured *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells. See legend of Figure 1 for further explanations. Average values of 3 experiments are shown for the viewpoints Eμ-HD (*Pax5*$^{-/-}$ *Rag2*$^{-/-}$), HS3B-HD (*Pax5*$^{-/-}$ *Rag2*$^{-/-}$) and 7183.16-HD (both genotypes). All other 4C-seq experiments were performed once.

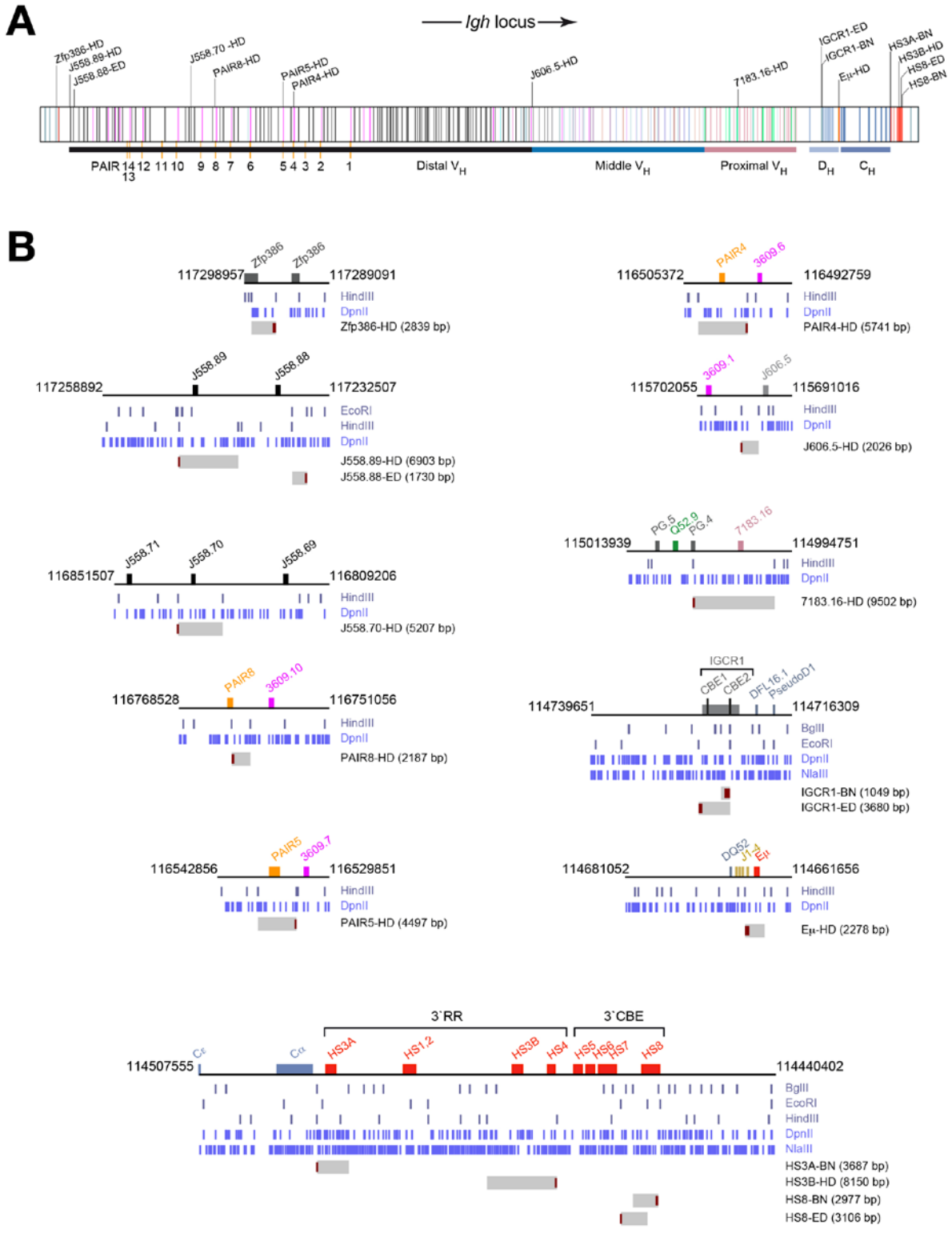
A
Igh locus
Zfp386-HD
J558.89-HD
J558.88-ED
J558.70 -HD
PAIR8-HD
PAIR5-HD
PAIR4-HD
J606.5-HD
7183.16-HD
IGCR1-ED
IGCR1-BN
Eμ-HD
HS3A-BN
HS3B-HD
HS8-ED
HS8-BN
PAIR 14 12 11 10 9 8 7 6 5 4 3 2 1
13
Distal V_H
Middle V_H
Proximal V_H
D_H
C_H
B
117298957
Zfp386
Zfp386
117289091
HindIII
DpnII
Zfp386-HD (2839 bp)
116505372
PAIR4
3609.6
116492759
HindIII
DpnII
PAIR4-HD (5741 bp)
117258892
J558.89
J558.88
117232507
EcoRI
HindIII
DpnII
J558.89-HD (6903 bp)
J558.88-ED (1730 bp)
115702055
3609.1
J606.5
115691016
HindIII
DpnII
J606.5-HD (2026 bp)
116851507
J558.71
J558.70
J558.69
116809206
HindIII
DpnII
J558.70-HD (5207 bp)
115013939
PG.5
Q52.9
PG.4
7183.16
114994751
HindIII
DpnII
7183.16-HD (9502 bp)
116768528
PAIR8
3609.10
116751056
HindIII
DpnII
PAIR8-HD (2187 bp)
114739651
IGCR1
CBE1
CBE2
DFL16.1
PseudoD1
114716309
BglII
EcoRI
DpnII
NlaIII
IGCR1-BN (1049 bp)
IGCR1-ED (3680 bp)
116542856
PAIR5
3609.7
116529851
HindIII
DpnII
PAIR5-HD (4497 bp)
114681052
DQ52
J1-4
Eμ
114661656
HindIII
DpnII
Eμ-HD (2278 bp)
3`RR
3`CBE
114507555
Cε
Cα
HS3A
HS1,2
HS3B
HS4
HS5
HS6
HS7
HS8
114440402
BglII
EcoRI
HindIII
DpnII
NlaIII
HS3A-BN (3687 bp)
HS3B-HD (8150 bp)
HS8-BN (2977 bp)
HS8-ED (3106 bp)

Figure S3

Figure S3. Description of viewpoints used for 4C-seq analysis.
(A) Schematic diagram of the mouse C57BL/6 *Igh* locus indicating the positions of the different viewpoints together with the location of PAIR, V_H, D_H and C_H elements. The distinct V_H gene families in the distal, middle and proximal V_H gene regions (Johnston et al., 2006) are indicated in different colors, and the Eμ enhancer and DHS sites in the 3' regulatory region are shown in red. (B) Description of viewpoints. Each viewpoint is referred to by the annotated feature present in or close to the primary restriction fragment (grey bar) and by the combination of primary and secondary restriction enzymes used for 4C template preparation. A brown bar at one end of the grey viewpoint fragment indicates the small HindIII-DpnII (HD), EcoRI-DpnII (ED) or BglII-NlaIII (BN) restriction fragment that was used for inverse PCR to generate the 4C-seq library. HS8 was previously referred to as site "38" by Garrett et al. (2005). The mm9 genomic sequence coordinates of mouse chromosome 12 are indicated for each region shown.

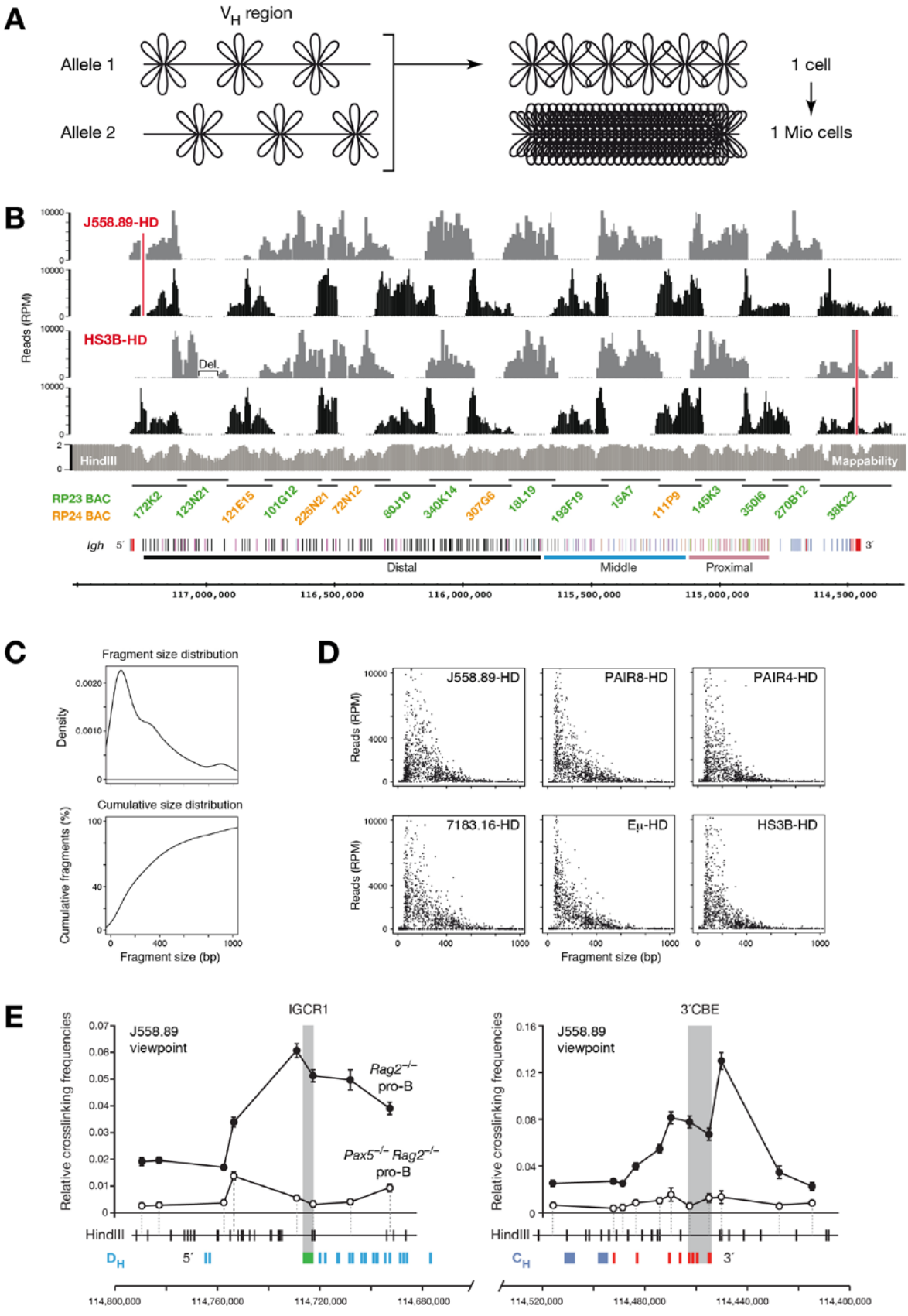
A
V_H region
Allele 1
Allele 2
1 cell
1 Mio cells
B
J558.89-HD
HS3B-HD
Del.
Reads (RPM)
HindIII
Mappability
RP23 BAC
RP24 BAC
172K2
123N21
121E15
101G12
226N21
72N12
80J10
340K14
307G6
18L19
193F19
15A7
111P9
145K3
350I6
270B12
38K22
Igh
Distal
Middle
Proximal
117,000,000
116,500,000
116,000,000
115,500,000
115,000,000
114,500,000
C
Fragment size distribution
Density
Cumulative size distribution
Cumulative fragments (%)
Fragment size (bp)
D
J558.89-HD
PAIR8-HD
PAIR4-HD
7183.16-HD
Eμ-HD
HS3B-HD
Reads (RPM)
Fragment size (bp)
E
IGCR1
3´CBE
J558.89 viewpoint
Rag2−/− pro-B
Pax5−/− Rag2−/− pro-B
Relative crosslinking frequencies
HindIII
D_H
C_H
5´
3´
114,800,000
114,760,000
114,720,000
114,680,000
114,520,000
114,480,000
114,440,000
114,400,000

Figure S4

Figure S4. Analysis of random BAC libraries.
(A) Schematic diagram indicating how flexible local interactions in the V_H gene region result in a continuum of chromatin loops across the V_H gene cluster in a population of 1 Mio cells. (B) Sequence pattern of random BAC libraries. The extent and number of each BAC are shown above the annotation of the *Igh* locus. For each viewpoint, two random BAC libraries were generated by using alternating BACs (indicated in grey or black). The sequencing pattern of each BAC was normalized by setting the highest peak to an RPKM value of 10000. The number of each RP23 or RP24 BAC is indicated in green or orange, respectively. An internal deletion (Del.) in BAC RP23-123N21 is indicated. (C) Frequency (density) and cumulative distribution of HindIII-DpnII fragments in the *Igh* locus according to increasing fragment size (in bp). (D) HindIII-DpnII fragment length bias in the PCR-amplified and combined random BAC libraries generated with each viewpoint shown. Each dot indicates the uniquely mapped sequence reads and fragment length for each HindIII-DpnII fragments of the *Igh* locus. (E) 3C-qPCR analysis of long-range interactions between the distal J558.89 region and HindIII fragments in the proximal IGCR1 (left) and 3'CBE (right) regions. HindIII digestion was used to prepare sixteen and three 3C-templates from *Rag2*$^{-/-}$ (filled circles) and *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ (open circles) pro-B cells on the C57BL/6 background, respectively. The crosslinking frequencies at the *Igh* and control *Ercc3* loci were calculated by using HindIII-digested and randomly ligated BAC DNA of these loci as a standard for PCR amplification. The relative crosslinking frequency was determined as the ratio of the crosslinking frequency at the indicated HindIII fragments of the *Igh* locus relative to the crosslinking frequency at the *Ercc3* gene, and the calculated value is shown in the middle of the respective HindIII fragment together with the standard error of the mean. These 3C-qPCR data demonstrate the specificity of the long-range interactions between the distal J558.89 sequences and the IGCR1 or 3'CBE region relative to the random BAC library control that was used for PCR normalization.

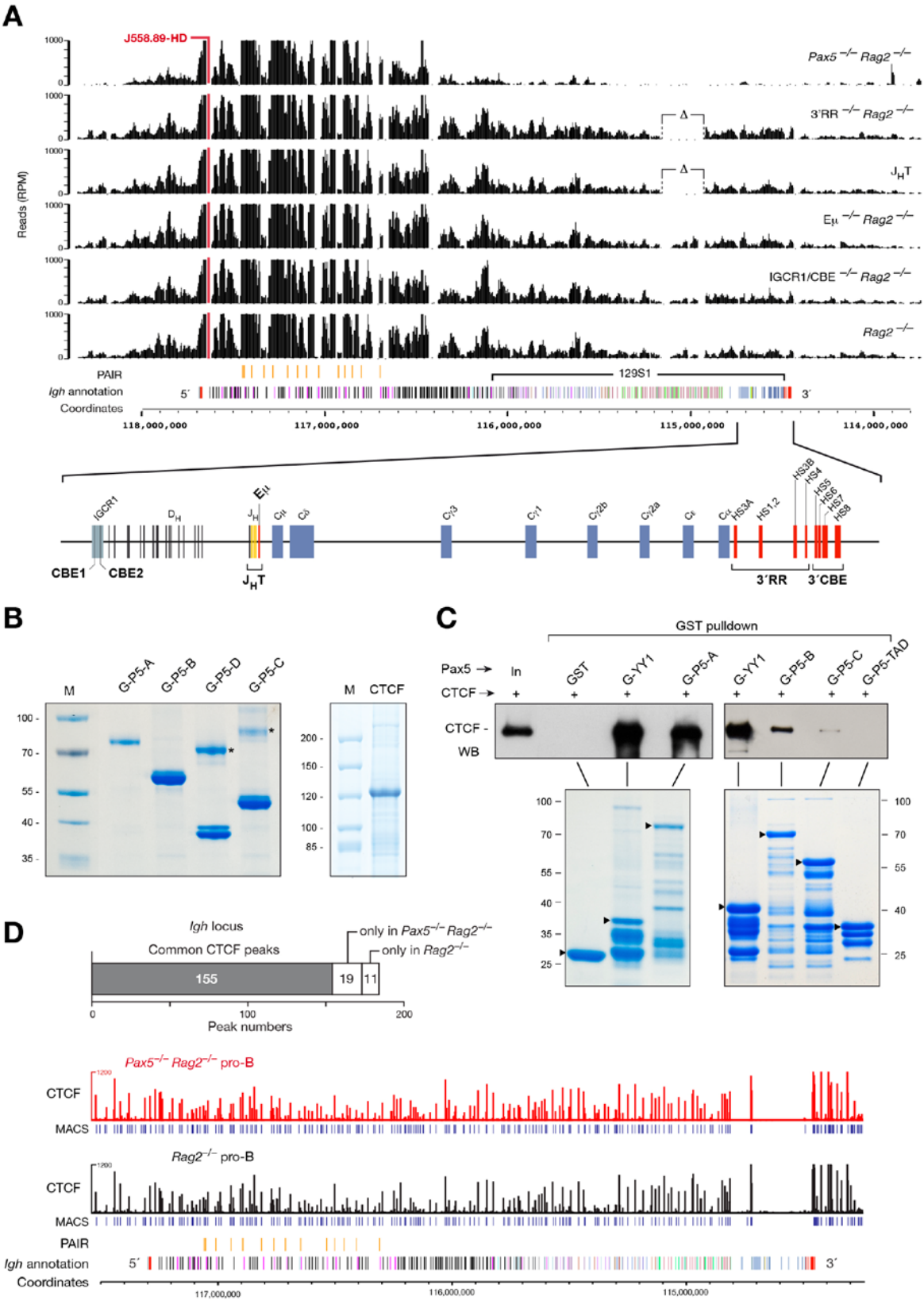
A
J558.89-HD
Pax5 −/− Rag2 −/−
3′RR −/− Rag2 −/−
Δ
JHT
Eμ −/− Rag2 −/−
IGCR1/CBE −/− Rag2 −/−
Rag2 −/−
Reads (RPM)
PAIR
129S1
Igh annotation
5′
3′
Coordinates
118,000,000
117,000,000
116,000,000
115,000,000
114,000,000
IGCR1
DH
JH
Eμ
Cμ
Cδ
Cγ3
Cγ1
Cγ2b
Cγ2a
Cε
Cα
HS3A
HS1,2
HS3B
HS4
HS5
HS6
HS7
HS8
CBE1
CBE2
JHT
3′RR
3′CBE
B
M
G-P5-A
G-P5-B
G-P5-D
G-P5-C
100
70
55
40
35
M
CTCF
200
150
120
100
85
C
GST pulldown
Pax5 →
CTCF →
In
GST
G-YY1
G-P5-A
G-YY1
G-P5-B
G-P5-C
G-P5-TAD
CTCF - WB
100
70
55
40
35
25
D
Igh locus
Common CTCF peaks
only in Pax5−/− Rag2−/−
only in Rag2−/−
155
19
11
0
100
200
Peak numbers
Pax5−/− Rag2−/− pro-B
CTCF
MACS
Rag2−/− pro-B
CTCF
MACS
PAIR
Igh annotation
5′
3′
Coordinates
117,000,000
116,000,000
115,000,000

Figure S5

Figure S5. Role of CTCF and 3' regulatory elements in long-range *Igh* interactions.
(A) 4C-seq analysis. The interaction pattern of the 129Sv *Igh* locus in short-term cultured pro-B cells of the indicated genotypes was determined by 4C-seq analysis with the viewpoint J558.89-HD (red). The J_HT pro-B cells were homozygous for the deletion eliminating the D_HQ52, J_H and Eμ elements (Gu et al., 1993). The 129Sv sequence reads were mapped to genomic C57BL/6 sequences except for the 3' half of the *Igh* locus, which was replaced by the known 1.6-Mb genomic sequences of the 129S1 *Igh* allele (Retter et al., 2007). A 120-kb deletion (Δ) was present in the E14 (129Ola) ES cells used to generate the J_HT and 3'RR mutant alleles. The positions of the analyzed regulatory elements in the 3' proximal domain are shown below the *Igh* annotation. (B) Analysis of purified CTCF and GST-Pax5 fusion proteins that were used for the GST-pulldown experiment shown in Figure 6F. The hexahistidine-tagged CTCF and the four GST-Pax5 fusion proteins (G-P5-A,B,C and D) were expressed in baculovirus-infected Sf9 cells and affinity-purified from cell lysates by Ni-NTA agarose chromatography or by precipitation with glutathione-sepharose beads, respectively. The purified proteins were analyzed by SDS-PAGE and visualized by Coomassie blue staining. Asterisks indicate protein dimers. M, protein size marker (in kilodaltons); G, GST; P5, Pax5. (C) GST-pulldown experiment with baculovirus-expressed CTCF and bacterially expressed GST fusion proteins. The CTCF protein in the precipitate was detected by Western blotting with CTCF antibodies (upper part). The purified proteins bound to the glutathione-sepharose beads were analyzed by SDS-PAGE (lower part). Arrows indicated the positions of the intact GST fusion proteins. Full-length YY1 sequences (amino acids 2-414) were present in the GST-YY1 (G-YY1) protein and the Pax5 transactivation domain (TAD; amino acids 304-366) in the G-P5-TAD protein. The GST-YY1 protein was used as a positive control, as YY1 was previously shown to interact with CTCF in GST-pulldown assays (Donohoe et al., 2007). Input (In; 1/20) is shown. (D) Similar CTCF binding pattern at the *Igh* locus in the presence or absence of Pax5. Short-term cultured *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells were analyzed by ChIP with a CTCF antibody (Millipore [07-729]) followed by paired-end Illumina sequencing resulting in a read length of 50 nucleotides. CTCF peaks were called by the MACS program (Zhang et al., 2008) with a p-value of $< 10^{-10}$. The majority (155) of CTCF peaks at the *Igh* locus were common to both cell types, whereas only 11 and 19 peaks were uniquely called in *Rag2*$^{-/-}$ and *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ pro-B cells, respectively.

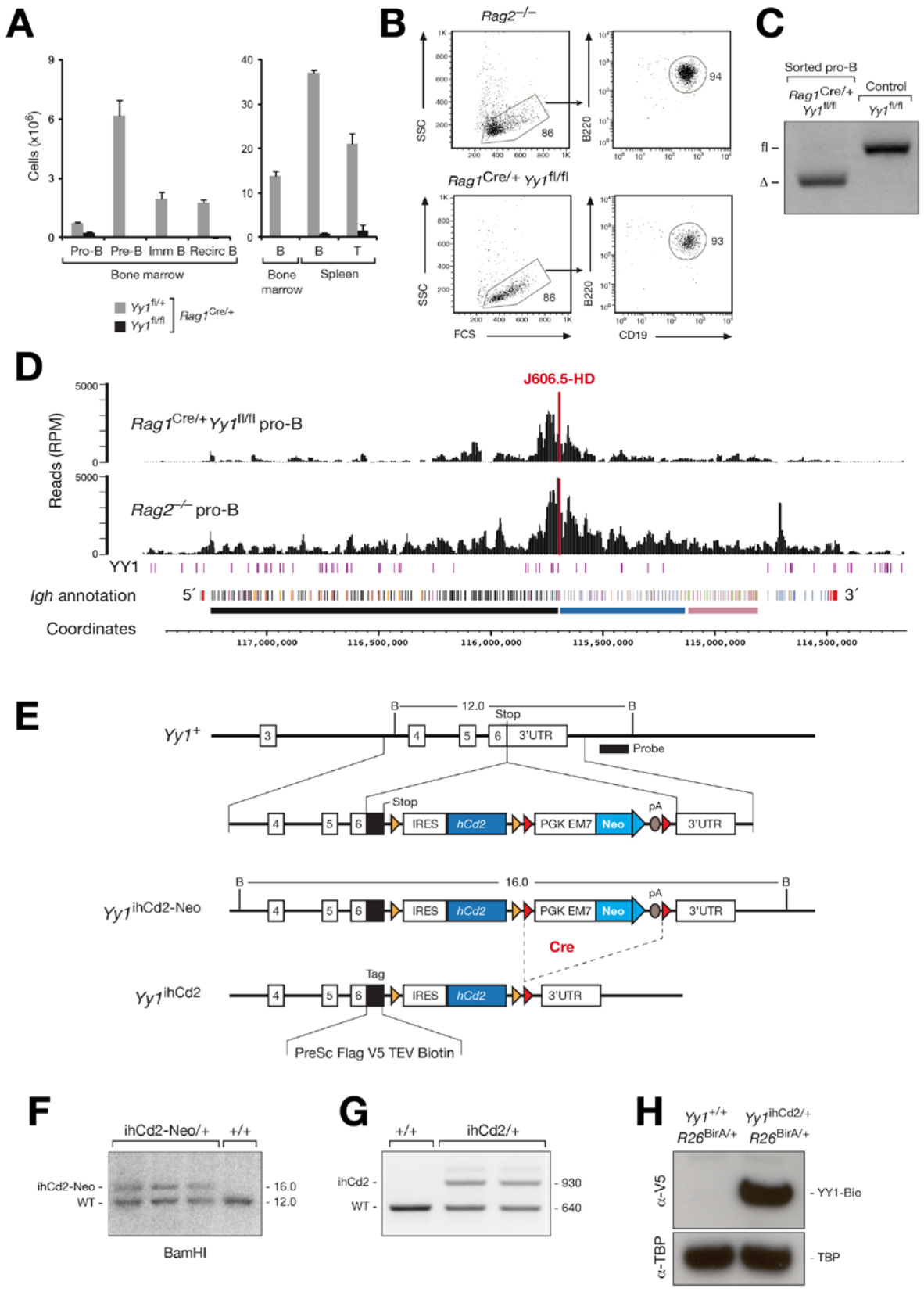
A
Cells (x10^6)
Pro-B
Pre-B
Imm B
Recirc B
Bone marrow
B
B
T
Bone marrow
Spleen
Yy1fl/+
Yy1fl/fl
Rag1Cre/+
B
Rag2−/−
SSC
B220
86
94
Rag1Cre/+ Yy1fl/fl
93
FCS
CD19
C
Sorted pro-B
Control
Rag1Cre/+ Yy1fl/fl
Yy1fl/fl
fl –
Δ –
D
J606.5-HD
Reads (RPM)
Rag1Cre/+Yy1fl/fl pro-B
Rag2−/− pro-B
YY1
Igh annotation
5´
3´
Coordinates
117,000,000
116,500,000
116,000,000
115,500,000
115,000,000
114,500,000
E
Yy1+
B
12.0
Stop
3'UTR
Probe
IRES
hCd2
PGK EM7
Neo
pA
Yy1ihCd2-Neo
16.0
Cre
Yy1ihCd2
Tag
PreSc Flag V5 TEV Biotin
F
ihCd2-Neo/+
+/+
ihCd2-Neo -
WT -
- 16.0
- 12.0
BamHI
G
+/+
ihCd2/+
ihCd2 -
WT -
- 930
- 640
H
Yy1+/+ R26BirA/+
Yy1ihCd2/+ R26BirA/+
α-V5
- YY1-Bio
α-TBP
- TBP

Figure S6

Figure S6. **4C-sequencing of YY1-deficient pro-B cells and characterization of the *Yy1*^{ihCd2} allele.** (A) Arrest of B cell development at the pro-B cell stage in *Rag1*$^{Cre/+}$ *Yy1*$^{fl/fl}$ mice. Bone marrow from control *Rag1*$^{Cre/+}$ *Yy1*$^{fl/+}$ (grey bars) and *Rag1*$^{Cre/+}$ *Yy1*$^{fl/fl}$ mice (black bars) at the age of 6 weeks were analyzed by flow cytometry, and the absolute cell numbers with the standard error of the mean are shown for pro-B cells (c-Kit^{+}CD19^{+}IgM^{-}IgD^{-}), pre-B cells (CD19^{+}CD25^{+}IgM^{-}IgD^{-}), immature (imm) B cells (CD19^{+}IgM^{+}IgD^{-}) and recirculating (recirc) B cells (CD19^{+}IgM^{-}IgD^{+}). Four mice were analyzed per genotype. (B) MACS sorting of YY1-deficient pro-B cells for 4C-seq analysis. Pro-B cells from the bone marrow of *Rag2*$^{-/-}$ and *Rag1*$^{Cre/+}$ *Yy1*$^{fl/fl}$ mice were stained with CD19-PE and MACS-sorted with PE-coupled MACS beads. The purity of the sorted pro-B cell populations was determined as relative percentage by flow cytometric analysis. (C) Cre-mediated deletion of the floxed *Yy1* allele in sorted pro-B cells. The MACS-sorted pro-B cells from *Rag1*$^{Cre/+}$ *Yy1*$^{fl/fl}$ mice (B) and control *Yy1*$^{fl/fl}$ pro-B cells were analyzed by PCR genotyping (Liu et al., 2007). The PCR fragments corresponding to the deleted (Δ) or intact (fl) floxed *Yy1* allele are indicated to the left of the gel. (D) Absence of long-range interactions at the *Igh* locus in YY1-deficient pro-B cells. MACS-sorted pro-B cells from the bone marrow of *Rag1*$^{Cre/+}$ *Yy1*$^{fl/fl}$ mice (B,C) and short-term cultured *Rag2*$^{-/-}$ pro-B cells were analyzed by 4C-sequencing with the viewpoint J606.5-HD. The YY1-binding sites (Figure 7B) at the *Igh* locus are shown. (D) Structure of the *Yy1*^{ihCd2} allele. A C-terminal tag sequence, an IRES-*hCd2* (ihCd2) reporter gene and a neomycin (Neo) resistance gene under the control of the mouse phosphoglycerate kinase (PGK) promoter were inserted between the last codon and the 3' untranslated region (3'UTR) of the *Yy1* gene. Brackets indicate the two homology regions mediating recombination in ES cells. The last *Yy1* exons are shown as open boxes. *LoxP* and *frt* sites are indicated by red and yellow arrowheads, respectively. The BamHI (B) fragments, which are indicative of the wild-type and *Yy1*$^{ihCd2\text{-}neo}$ alleles, are shown together with their length (in kb). The tag sequences added at the last *Yy1* codon contained cleavage sites for the PreScission (PreSc) and TEV proteases, epitopes for Flag and V5 antibodies and a biotin acceptor sequence (Biotin) for biotinylation by the *Escherichia coli* biotin ligase BirA (de Boer et al., 2003). pA, polyadenylation site. (F) Southern blot analysis of correctly targeted ES cell clones. BamHI-digested DNA was analyzed by hybridization with the DNA probe shown in (E). (G) PCR genotyping of tail DNA isolated from *Yy1*$^{ihCd2/+}$ and *Yy1*$^{+/+}$ mice. (F) Efficient expression of the YY1-Bio protein in pro-B cells of *Yy1*$^{ihCd2/+}$ *Rosa26*$^{BirA/+}$ mice. Nuclear lysates of *Yy1*$^{ihCd2/+}$ *Rosa26*$^{BirA/+}$ and *Rosa26*$^{BirA/+}$ pro-B cells were analyzed by immunoblotting with anti-V5 and anti-TBP antibodies.

Table S1: 4C-sequencing primers

Illumina adaptors:

Sequence S1 5'-AATGATACGGCGACCACCGA..
..ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
Sequence S2 5'-CAAGCAGAAGACGGCATACGA-3'

4C-seq primers:

Zfp386-HD	HindIII	5'-S1-GAAAAGACAAAAGTGA**AAGCT**-3'
	DpnII	5'-S2-ATTACCTTTGTGACTTGTGGA-3'
J558.89-HD	HindIII	5'-S1-AAAGGCATGTTAGCT**AAGCTT**-3'
	DpnII	5'-S2-CATTCTCCATCCATTAGGAT-3'
J558.88-ED	EcoRI	5'-S1-TTCCAGAACCTTCCTACCT-3'
	DpnII	5'-S2-ATCACATGTTAATCACTGCAG-3'
J558.70-HD	HindIII	5'-S1-GGCATGCTAGCT**AAGCTT**-3'
	DpnII	5'-S2-ATTCTCCATGCATTAGGATC-3'
PAIR8-HD	HindIII	5'-S1-TGGTATATGGAAAGG**AAGC**-3'
	DpnII	5'-S2-CGACACCAAACCTAATCAAG-3'
PAIR5-HD	HindIII	5'-S1-AATCAGTCAGCTACAC**AAGC**-3'
	DpnII	5'-S2-CGACACCAAACCTAATCAAG-3'
PAIR4-HD	HindIII	5'-S1-TGTAGTGGAACAATACTCT**AAGCTT**-3'
	DpnII	5'-S2-CTTAGAGAGAATAAGATTTGTCTAATGT-3'
J606.5-HD	HindIII	5'-S1-AGTTCAGGACATGTGTAAGTT**AA**-3'
	DpnII	5'-S2-ATCCATTAATGCCATTGTTC-3'
7183.16-HD	HindIII	5'-S1-TCCCCTAGACTTCTC**AAGC**-3'
	DpnII	5'-S2-GTTGGACCTAACATATTCAGG-3'
IGCR1-ED	EcoRI	5'-S1-CACAGGTACCCAGATA**GAATTC**-3'
	DpnII	5'-S2-GTGGTTGGAGGGTTGATC-3'
IGCR1-BN	BglII	5'-S1-ACAGCAGCAGGG**AGATCT**-3'
	NlaIII	5'-S2-TCTGTAGATTGACTGGAGATTT-3'
Eμ-HD	HindII	5'-S1-TGCCTCAGACTTC**AAGCTT**-3'
	DpnII	5'-S2-GGGAAATAAACTGTCTAGGGAT-3'
HS3A-BN	BglII	5'-S1-TTTAGAAATCAGGTCC**AGATC**-3'
	NlaIII	5'-S2-CTATATTGGTAACTGAACCGTG-3'
HS3B-HD	HindIII	5'-S1-CCCACCTAACTCC**AAGCTT**-3'
	DpnII	5'-S2-CAGACATGTGGGCTGAGAT-3'
HS8-ED	EcoRI	5'-S1-GACACTGGCCCTCA**GAAT**-3'
	DpnII	5'-S2-TGGAACAGTGGAGTCAGAT-3'
HS8-BN	BglII	5'-S1-GCCTATATCCCTGTGACCAC-3'
	NlaIII	5'-S2-GCCAGTGTTGACTGGACAC-3'

Table S2: Primers used for 3C-qPCR experiments

Primers (5' to 3')	**Description**	**Figure**
TAGGGATAACAGCGCAATCC	Genomic DNA quant. (*Rnr2*) *1	Fig. 6C, S4E
GACTTTAATCGTTGAACAAACGAAC	Genomic DNA quant. (*Rnr2*) *1	Fig. 6C, S4E
GCCCTCCCTGAAAATAAGGA	Positive control (*Ercc3*) *2	Fig. 6C, S4E
GACTTCTCACCTGGGCCTACA	Positive control (*Ercc3*) *2	Fig. 6C, S4E
AAAATTTCTGTGGCATCATTGCT	IGCR1 (viewpoint)	Fig. 6C
CTGAATGTAGTGGAACAATACTCT	PAIR4	Fig. 6C
AAGGAACTTGATTTATTTTGGCAT	PAIR8	Fig. 6C
TATTCCTGCAAAGGCATGTTAGC	J558.89	Fig. 6C
TGTGTGCTTAGAAAGTTTTACATCT	J558.89 (viewpoint)	Fig. S4E
ATCACTGGTATTCTTGCTCC	Chr12: 114787053-114791569	Fig. S4E
CAAAACCCAATGACCAGTCCA	Chr12: 114778143-114787053	Fig. S4E
TGGTTAGCATCAGAAACCCCTT	Chr12: 114755043-114759774	Fig. S4E
TTCCACTGGCTCAAGAAGAACTGACA	Chr12: 114751970-114755043	Fig. S4E
GAATCACTCAGCAGCCCTA	Chr12: 114723014-114734838	Fig. S4E
CTCTATTTCTACTTTCTCCAAGC	Chr12: 114722112-114723014	Fig. S4E
ATTGGGCAACACTCTTAATGTGA	Chr12: 114693994-114722112	Fig. S4E
CTCAGTTCCTGCAACATGCTC	Chr12: 114691294-114693994	Fig. S4E
AACTGTGTGCTAGACGGACA	Chr12: 114686568-114691294	Fig. S4E
TGGACCTGTATCAAAACGGTGCTC	Chr12: 114510165-114522182	Fig. S4E
GAGGCCAAATGTTTAGAAATCAGGTCCA	Chr12: 114491123-114493852	Fig. S4E
TGGCCTTTGACTCCGTTT	Chr12: 114486613-114491123	Fig. S4E
AGGTGTTAAGGAAAACTTGCTC	Chr12: 114480874-114486613	Fig. S4E
GCAGCTCACCTCTGTAATCCCA	Chr12: 114474053-114474923	Fig. S4E
GCCTGCACTCACCTACCCA	Chr12: 114465909-114474053	Fig. S4E
CAGGGGCTCTGCATCATACCAA	Chr12: 114459545-114465909	Fig. S4E
CGGTGACTCTCTTTTTGTCTGTCC	Chr12: 114450785-114459545	Fig. S4E
CCCTTCAAACTGTACTAGCAGG	Chr12: 114449888-114450785	Fig. S4E
ACTCTCCCTGCTGATAGACA	Chr12: 114420753-114434637	Fig. S4E
CTGGTGTCCTTCCGCCTT	Chr12: 114409003-114420753	Fig. S4E

Taqman probe

CGACCTCGATGTTGGATCAGGACATCC	Genomic DNA quant. (*Rnr2*) *1	Fig. 6C, S4E
AAAGCTTGCACCCTGCTTTAGTGGCC	Positive control (*Ercc3*) *2	Fig. 6C, S4E
ACT[+G]CT[+C]TG[+A]GC[+C]TGTCCT	IGCR1 *3	Fig. 6C
CCT[+G]CA[+A]AG[+G]CA[+T]GTTAGC	J558.89 *3	Fig. S4E

*1: Lefever et al. (2009) Nucleic Acids Res. *37*, D942-D945.

*2: Splinter et al. (2011) Genes Dev. *25*, 1371-1383.

*3: Nucleotides with [+N] are locked nucleic acid residues (LNA®)

The sequence coordinates correspond to the respective mm9 genomic sequences (C57BL/6).

Supplemental Experimental Procedures

Mice

The *Pax5*$^{+/-}$ (Urbánek et al., 1994), *Pax5*$^{Bio/Bio}$ (McManus et al., 2011), J_HT (Gu et al., 1993), *Yy1*$^{fl/fl}$ (Liu et al., 2007), *Rag1*$^{Cre/+}$ (McCormack et al., 2003) and *Rosa26*$^{BirA/BirA}$ (Driegen et al., 2005) mice were maintained on the C57BL/6 background, the *Rag2*$^{-/-}$ (Shinkai et al., 1992) on both the C57BL/6 and 129Sv background and the IGCR1/CBE mutant (*Igh*$^{tm5.1Alt}$; Guo et al., 2011), Eμ mutant (Perlot et al., 2005) and 3'RR mutant (Vincent-Fabert et al., 2010) mice on the 129Sv background. As the *Yy1* gene, which was originally targeted in 129Sv ES cells (Liu et al., 2007), is separated by only 4.4 megabases from the 3' end of the *Igh* locus on mouse chromosome 12, the *Igh* sequences are still of 129Sv origin despite backcrossing of the floxed (fl) *Yy1* allele into the C57BL/6 background. All animal experiments were carried out according to valid project licenses, which were approved and regularly controlled by the Austrian Veterinary Authorities.

Generation of the *Yy1*^{ihCd2} allele

The targeting vector for generating the *Yy1*^{ihCd2} allele was obtained by first inserting the following sequences in the 5' to 3' direction (see Figure S6E) into the *Yy1* BAC RP23-351K4 by recombineering in *E. coli*; 233-bp MfeI-AscI fragment containing the C-terminal tag sequences (fused in frame to the last *Yy1* codon), 1.8-kb AscI-SalI fragment containing the IRES-*hCd2* (ihCd2) reporter gene (flanked by *frt* sites) and 1.9-kb SalI-XhoI fragment containing the mouse phosphoglycerate kinase promoter linked to the neomycin resistance gene (flanked by *loxP* sites). The tag sequences contained cleavage sites for the PreScission (PreSc) and TEV proteases, epitopes for Flag and V5 antibodies and a biotin acceptor sequence (Biotin) for biotinylation by the *Escherichia coli* biotin ligase BirA (de Boer et al., 2003). In a second step, the *Yy1*^{ihCd2} targeting vector was generated by excising and inserting the integrated sequences together with the flanking 5' (4.9 kb) and 3' (1.9 kb) homology regions by recombineering from the modified BAC into the pBV-DTA-pA plasmid containing an *HSV-TK* gene (for negative selection). SgrA1-linearized DNA (15 μg) was electroporated into cells ($1x10^7$) of the hybrid C57BL/6 x 129Sv ES cell line A9 followed by selection with 250 μg/ml G418. PCR-positive clones were verified by Southern blot analysis of BamHI-digested DNA (Figure S6F) prior to injection into C57BL/6 blastocysts and the generation of *Yy1*$^{ihCd2\text{-}Neo/+}$ mice. The *Yy1*^{ihCd2} allele was obtained by crossing *Yy1*$^{ihCd2\text{-}Neo/+}$ mice with the *Mox2*-Cre line (Tallquist and Soriano, 2000). The following primers were used for PCR genotyping of *Yy1*$^{ihCd2/+}$ mice: (a) 5'-ATGCATGTAACTGCCCATTC-3'; (b) 5'-CACATGTGCGAATCCATACC-3'; (c) 5'-AGTAACGCCATTTTGCAAGG-3'. The wild-type *Yy1* allele was identified as a 640-bp PCR fragment with the primer pair a/b and the *Yy1*^{ihCd2} allele as a 930-bp PCR fragment with the primer pair a/c (Figure S6G). The *Yy1*$^{ihCd2/+}$ mouse was backcrossed into the C57BL/6 background.

FACS analysis and sorting

The following antibodies were used for flow cytometric analysis of bone marrow, spleen and thymus from 3-4-week-old mice: anti-B220/CD45R (RA3-6B2), CD3ε (145-2C11), CD4 (GK1.5), CD8a (53-6.7), CD11b/Mac1 (M1/70), CD19 (1D3), CD25/IL-2Rα (PC61), CD90.2/Thy1.2 (53-2.1), CD117/c-Kit (2B3), Gr1 (RB6-8C5), IgD (1.19), IgM^a (Igh-6a/DS-1), IgM^b (AF6-78), IgM (11/41) and TCRβ (H57-597).

In vitro culture of lymphoid progenitors

Rag2$^{-/-}$ and *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ pro-B cells were cultured on OP9 cells in IL-7 containing IMDM as described (Nutt et al., 1997).

Antibodies

The following antibodies were used: anti-CTCF (affinity-purified rabbit Ab; Millipore [07-729]) for co-immunoprecipitation and ChIP-seq analyses as well as biotinylated anti-Pax5 (rat mAb clone 1H9; from S. Nutt, WEHI Melbourne), anti-CTCF (mouse mAb clone 48, BD Transduction Laboratories [612148]), anti-V5 (mouse mAb clone 5C5; from F. Grosveld, Rotterdam), anti-GST (goat polyclonal Ab; GE Healthcare [PRN1236]) and anti-TBP (mouse mAb clone 3G3; from L. Tora, Illkirch) for immunoblot analysis.

Nuclear extract preparation and coprecipitation analysis

Nuclear extracts were prepared as described previously (Dyer and Herzog, 1995) with minor modifications. Briefly, 1-2x10^8 pro-B cells were lysed in a buffer consisting of 10 mM Tris pH 8.0, 0.32 M sucrose, 50 mM KCl, 20 mM NaCl, 3 mM $CaCl_2$, 2 mM magnesium acetate (MgAc), 0.1% NP-40, 1 mM DTT (omitted for extracts analyzed by immunoprecipitation), 2 mM 6-aminocaproic acid (6AA), 0.15 mM spermine, 0.5 mM spermidine, 0.5 μM MSF and 0.1% protease inhibitor cocktail, and the nuclei were collected by centrifugation for 5 minutes at 500 xg. Pelleted nuclei were then lysed in a lysis buffer consisting of 10 mM HEPES pH 7.9, 20% glycerol, 2.0 mM MgAc, 0.36 M KCl, 0.2% CHAPS, 10 mM NaF, 0.5 mM DTT, 2 mM 6-aminocaproic acid (6AA), 0.5 μM PMSF and 0.1% protease inhibitor cocktail. Where indicated in Figure 6E, 1.8 μl of the nuclease Benzonase^R (250 units/μl; Merck) was immediately added to 800 μl of the nuclei suspension (corresponding to 2x10^8 cells) followed by incubation at 4^{o}C for 45 min and centrifugation at 16,100 xg for 10 min. The protein content of the nuclear extract was measured to be ~15 mg/ml by Bradford assay (BioRad) using BSA as a standard. The extract was diluted with 3 volumes of the above lysis buffer containing 150 mM KCl prior to streptavidin pulldown or immunoprecipitation.

Dynal beads were blocked with 2 mg/ml BSA in PBS for 2 hr at room temperature. Nuclear extracts were incubated with streptavidin Dynal beads (M280; Invitrogen) or with purified antibodies at 4^{o}C overnight. In the latter case, protein A Dynal beads (Invitrogen) were subsequently added for another 3 hr. The beads were washed 5 times in 20 mM Tris pH 8, 250 mM KCl, 2 mM MgCl2, 10% glycerol, 1 mM DTT (omitted for immunoprecipitation), 2 mM 6-aminocapronic acid, 10 mM NaF, 10 mM β-glycerophosphate, 1 mM sodium pyrophosphate and 0.1% protease inhibitor cocktail followed by collecting the supernatant by magnetic sorting. The precipitated proteins were resuspended in 2x SDS sample buffer, eluted from the beads by boiling and separated by SDS-PAGE followed by immunoblotting.

In vitro binding assays with purified CTCF and Pax5 proteins

The GST-Pax5 fusion proteins, which were expressed from bacterial and baculovirus expression vectors, contained the following human Pax5 sequences; G-P5-A (Pax5 aa 1-391), G-P5-B (Pax5 aa 144-391) G-P5-C (Pax5 aa 144-301), G-P5-D (Pax5 aa 302-391) and G-P5-TAD (Pax5 aa 304-366). The GST-YY1 (G-YY1) protein contained the full-length mouse YY1 sequences (aa 2-414). A full-length human CTCF protein containing a N-terminal tag of six histidine residues was expressed from a baculovirus in infected Sf9 (*Spodoptera*

frugiperda) cells and was purified from cell lysates by Ni-NTA agarose affinity chromatography.

GST-Pax5, GST-YY1 and control GST proteins were expressed in the *E.coli* Rosetta^{TM}(DE3)pLysS strain (Novagen). Overnight cultures of bacteria containing the expression plasmids were diluted 1:100 in LB-medium containing 1% glucose and grown at 37°C to an absorbance $A_{600\ nm}$ of 0.6 before protein expression was induced by the addition of 0.2 mM isopropyl-β-thiogalactopyranoside for 4 h at 37°C. All buffers used in the lysis and affinity-binding reactions were supplemented with 2 mM DTT, 2 mM 6-aminocapronic acid, 10 mM NaF, 1 mM sodium pyrophosphate and 0.1% protease inhibitor cocktail. Bacteria were lysed in RIPA buffer (150 mM NaCl, 50 mM Tris pH 8, 0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate, 10% glycerol) supplemented with 0.2 mg/ml lysozyme (Sigma) and 0.1 U/ml of the nuclease Benzonase^{R} (Merck). After incubation on ice for 15 min, the lysate was diluted with the same volume of buffer A consisting of 350 mM KCl, 50 mM Tris pH 8, 0.1% SDS, 0.5% NP-40, 0.5% Triton X-100 and was incubated for additional 15 min at 4°C. After removal of cellular debris by centrifugation, the lysates were incubated with glutathione-sepharose 4B beads (GE Healthcare) for 3 h at 4°C. The sepharose beads were washed 4 times with buffer A followed by 2 washes with binding buffer B consisting of 50 mM NaCl, 150 mM KCl, 50 mM Tris pH 8, 2 mM $MgCl_2$, 0.5 mM $ZnCl_2$ and 10% glycerol. The amount of the GST-fusion proteins bound to the glutathione-sepharose beads was estimated by SDS-PAGE and Coomassie staining relative to a BSA standard. Approximately 25 μg of GST-Pax5, GST-YY1 and GST proteins purified from *E. coli* were incubated with 2.5 μg of recombinant CTCF for 3 h at 4°C in 400 μl of binding buffer B. After washing 4 times with buffer B, the bound proteins were resuspended in 2x SDS sample buffer, eluted from the beads by boiling and separated by SDS-PAGE followed by immunoblotting (Figure S5C).

In addition, the GST-Pax5 fusion proteins and a control GST-GFP (G-GFP) protein were expressed in baculoviruses-infected Sf9 cells. The proteins were purified from the cell lysates with glutathione-sepharose bead followed by elution with 10 mM reduced glutathione and subsequent purification on a Superdex 200 PC 3.2/30 column. The concentration of the purified proteins was measured with a Nanodrop spectrophotometer. To assay the Pax5-CTCF interaction, 5 μg of purified G-P5-B, C or D protein was bound to glutathione-sepharose beads for 2 h followed by the addition of 2.5 μg of purified recombinant CTCF in 400 μl of binding buffer C consisted of 250 mM NaCl, 50 mM Tris pH 8, 0.5 mM $ZnCl_2$ and 10 % glycerol (Figure 6F, right part). After incubation for 3 h at 4°C, the sepharose beads were washed 3 times with buffer D containing 120 mM KCl, 20 mM Tris pH 8, 2 mM $MgCl_2$ and 0.5 mM $ZnCl_2$. As the G-P5-A protein was expressed at a considerably lower level in Sf9 cells, we performed the binding assays with the G-P5-A and G-P5-B proteins bound to the glutathione-sepharose beads directly after GST-pulldown from the cell lysates without further purification on a Superdex 200 PC 3.2/30 column (Figure 6F, left part). Approximately 10 μg of G-P5-A and G-P5-B protein (quantified by Coomassie staining) were incubated with 2.5 μg of CTCF for 3 h at 4°C in 400 μl of binding buffer C followed by 3 washes of the sepharose beads with buffer D. The bound proteins of both experiments were resuspended in 2x SDS sample buffer, eluted from the beads by boiling and separated by SDS-PAGE followed by immunoblotting with an affinity-purified rabbit anti-CTCF Ab (Millipore [07-729]; Figure 6F, left part).

General description of 4C-sequencing experiments

To determine the 3D architecture of the *Igh* locus at high resolution, we have used the 4C-sequencing method, which facilitates the unbiased identification of all genomic sequences that contact a selected genomic site (referred to as 'viewpoint') within the 3D nuclear space in a given cell population (van de Werken et al., 2012). To this end, pro-B cells were fixed with formaldehyde, and the chromatin was digested with the 6 bp-recognizing restriction enzymes HindIII (H), EcoRI (E) or BglII (B) followed by ligation in dilution, which favor 'intramolecular' ligation between crosslinked DNA fragments (Dekker et al., 2002). After reversal of the crosslinks, the ligated DNA was digested with the 4-bp cutter DpnII (D) or NlaIII (N) and religated in dilution to create small DNA circles, which were used as template for inverse PCR with viewpoint-specific primers to amplify all DNA sequences contacting the chromosomal site of the viewpoint (Simonis et al., 2007). The PCR fragments were quantified by deep sequencing, and the association frequencies of the interacting DNA fragments were determined as 20-kb running mean values of the 4C-seq reads normalized per million mapped sequence reads (RPM; van de Werken et al., 2012). For our study, we have used 16 different viewpoints across the *Igh* locus (Figure S3A), which are referred to by the annotated feature present in or close to the primary restriction fragment and by the combination of the primary and secondary restriction enzymes used (Figure S3B). Moreover, as genomic fragments in the *Igh* locus can also be brought into close proximity through deletion of the intervening sequences by V(D)J rearrangements, only pro-B cells on the recombination-deficient *Rag2* mutant background (Shinkai et al., 1992) were analyzed by 4C-sequencing.

Detailed description of 4C-sequencing experiments

Pro-B cells isolated from the bone marrow of *Rag2*$^{-/-}$, *Rag2*$^{-/-}$ Eμ mutant, *Rag2*$^{-/-}$ 3'RR mutant, *Rag2*$^{-/-}$ IGCR1/CBE mutant and homozygous J_HT mice were expanded in vitro for 4 days on OP9 cells in the presence of IL-7. Pro-B cells from the bone marrow of *Rag2*$^{-/-}$ and *Rag1*$^{Cre/+}$ *Yy1*$^{fl/fl}$ mice were stained with CD19-PE antibody and anti-PE MACS beads prior to MACS-sorting, resulting in a purity of 94% (Figure S6B). Total thymocytes were isolated from wild-type C57BL/6J mice by passing the thymi through cell strainers and removing the dead cells by density centrifugation (Lympholyte M, Cedarlane Lab. [CEDCL5031]).

The 4C templates were prepared as previously described (Simonis et al., 2007) with the following modifications. For sequestration of SDS, 50 μl of 20% Triton-X 100 was incubated with the chromatin for 2-3 hours. The total amount of the primary restriction enzyme (HindIII, EcoRI or BglII) equaled 700 U and was added as follows: 300 U overnight, 200 U for 4 hours and finally 200 U for another 4-6 hours. The digested chromatin was ligated overnight at 16^{o}C followed by a 1 hr incubation at room temperature. In the second digestion round with the secondary restriction enzyme, the ligation products were incubated with 50 U of NlaIII (NEB, R0125S) overnight or 150 U of DpnII (NEB, R0543T), which were added sequentially (50 U) prior to overnight ligation at 16^{o}C as described above. After ethanol precipitation, the final 4C template was purified twice using the Wizard® SV Gel and PCR Clean-Up System (Promega, A9282), each time with one column per sample prepared from 1x10^{7} cells. The DNA concentration was measured with the Nanodrop spectrophotometer. Approximately 20 μg of 4C template was obtained from 1x10^{7} pro-B cells. For each viewpoint, 16 individual PCR amplifications were performed with ~150 ng of 4C template by using specific primers that were extended with Illumina sequencing adaptors (Table 1). Thus, ~2.4 μg of 4C template corresponding to approximately 1x10^{6} pro-B cells were analyzed for

each viewpoint. The PCR amplification was performed with 35 cycles, as too little PCR product was generated after 30 cycles. The PCR products of pooled reactions were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). After column purification, the PCR products obtained with the third *Rag2*$^{-/-}$ and *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ replica 4C templates as well as the *Rag2*$^{-/-}$ Eμ mutant, *Rag2*$^{-/-}$ 3'RR mutant, *Rag2*$^{-/-}$ IGCR1/CBE mutant and *Rag1*$^{Cre/+}$ *Yy1*$^{fl/fl}$ 4C templates were selected for a fragment size below 1 kb as follows**.** Ten percent of the purified and pooled reactions were quantified on an agarose gel by measuring ethidium bromide incorporation. PCR products from cells of the same genotype were pooled in ratios, which yielded equal amounts of amplified DNA between 100 and 1000 bp. The pooled DNA was separated on a 2% agarose gel (Lonza [Nusieve GTG agarose, 50080]). After cutting out the lanes, the agarose was degraded with β-agarase I enzyme (NEB, M0392L). The DNA was extracted twice with phenol and precipitated with ethanol after phenol removal with chloroform. The precipitate was dissolved in 10 mM Tris pH 8 and quantified by the Bioanalyzer instrument. Cluster generation and single-end sequencing was carried out by using the Illumina GAIIx and HiSeq 2000 systems according to the manufacturer's guidelines, which resulted in a read length of 76 (GAIIx) or 100 (HiSeq 2000) nucleotides.

4C-sequencing of 129Sv pro-B cells with regulatory mutations in the *Igh* 3' proximal region

Pro-B cells carrying the dual CBE mutation of the IGCR1 element (Guo et al., 2011), the core Eμ enhancer deletion (Perlot et al., 2005) or a deletion of the entire 3'RR (Vincent-Fabert et al., 2010) were analyzed on the 129Sv *Rag2*$^{-/-}$ background (Figures 6A and S5A). We furthermore performed a 4C-seq experiment with pro-B cells containing the homozygous J_HT mutation, which prevents *Igh* recombination due to deletion of the D_HQ52, J_H and Eμ enhancer elements (Gu et al., 1993). As all four mutations were introduced into the *Igh* allele of the 129Sv mouse strain, we also performed control 4C-seq experiments with cultured *Pax5*$^{-/-}$ *Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ pro-B cells on the 129Sv background. Due to the sequence divergence of the *Igh* locus between the C57BL/6 and 129Sv mouse strains (Johnston et al., 2006; Retter et al., 2007), the 129Sv sequence reads could be accurately mapped only to the known 1.6-Mb genomic sequences at the 3' half of the 129S1 *Igh* locus (Retter et al., 2007), whereas reads in the distal V_H gene region were aligned to the C57BL/6 *Igh* sequences, thus resulting in relatively wide gaps (Figures 6A and S5A).

For mapping the 4C-seq reads obtained with 129Sv *Rag2*$^{-/-}$ pro-B cells, we generate the following mouse genomic sequence file. The mm9 sequences of the *Igh* locus (strain C57BL/6, chr 12: 114,000,000-117,500,000) were aligned with the reverse complement of the AJ851868 sequence (strain 129S1; Retter et al., 2007) using the Mauve 2.3.1 program (Darling et al., 2004) to map the 5' and 3' ends of the AJ851868 sequence within the mm9 *Igh* sequences. The C57BL/6 sequences between 114,493,850 and 115,696.909 were then replaced with the reverse complement of the AJ851868 sequence.

Random 4C-seq BAC library

The templates for random 4C-seq BAC libraries were generated using DNA mixtures of 8 to 9 C57BL/6J BAC clones (listed in Figure S4B), which contained the viewpoint of interest and covered the mouse *Igh* locus with every other BAC to avoid sequence overlaps (Figure S4B) that would have otherwise resulted in distortion of the sequence pattern at the junctions. BAC

DNA was mixed at equimolar ratio, and 25 μg of DNA was digested with 250 U HindIII in 250 μl volume overnight at 37°C followed by ligation with 62.5 U T4 DNA ligase in a 150 μl volume. The ligated DNA was digested by 120 U Sau3AI in a 250 μl volume. To favor intramolecular ligation, the second ligation was performed in dilution by ligating the Sau3AI-digested DNA in 4 ml (at 6.3 ng/μl) with 62.5 U T4 DNA ligase. Between each step, DNA was purified by phenol-chloroform extraction followed by ethanol precipitation. After the second ligation, DNA was purified by phenol-chloroform extraction, ehanol precipitation and the Wizard® SV Gel and PCR Clean-Up System (Promega, A9282). The BAC DNA template was mixed with a mouse genomic DNA template (prepared in the same way as the BAC template) at a ratio of 1:20 and was subjected to PCR amplification with viewpoint-specific primers. PCR amplification and subsequent DNA purification were performed as described in the 4C-sequencing section.

Bioinformatic analysis of 4C-seq data

The sequence reads of different 4C-seq experiments were demultiplexed by splitting the reads according to the 5' adapter sequence including the recognition sequence of the primary restriction enzyme by allowing up to 3 mismatches. For all experiments, the split reads were subsequently trimmed to 55 nucleotides (extending from the restriction enzyme site to the 3' end of the sequence. The trimmed reads were uniquely aligned to the mouse genome assembly version of July 2007 (NCBI37/mm9) using the Bowtie program version 12.5 (Langmead et al., 2009), allowing up to 2 mismatches (-v 2 --best --strata --tryhard -m 1 --chunkmbs 256). A genome-wide map of the pimary restriction enzyme was created using the program Emboss restrict, and only reads mapping to sequences at the restriction sites were retained. The depth of the reads at each restriction site was counted separately for each strand. The read count for each primary restriction fragment was determined by summing up the reads that were mapped at the 5' and 3' end of the restriction fragment.

The reads originating from self ligation of the viewpoint fragment were eliminated as follows. The restriction fragment of the viewpoint and 2 fragments on either side of the viewpoint were excluded from the analysis by setting the read count of these fragments to 0. Alternatively, the read counts present within 12 kb on either site were eliminated, if the combined length of the 2 fragments was smaller than 12 kb. After exclusion of the self-ligation product and the two adjacent restriction fragments, normalized read counts were calculated for each restriction fragment as *r*eads *p*er *m*illion of mapped sequence reads (RPMs). For visualization of the 4C-seq data, a running mean was determined by calculating the sum of the sequence counts in a window of 20 kb, divided by the number of fragments within the 20 kb window. The sliding window was moved with a step size of 5 kb.

Calculation of sequence mappability

A map containing only the mappable restriction sites of the genome was created by extracting the 55 nucleotide sequence on both sides of each restriction site and aligning these sequences to the mouse mm9 genomic sequences using the Bowtie program version 12.5 without allowing any mismatches (-v 0 --best --strata --tryhard -m 1 --chunkmbs 256). End sequences recovered by this procedure were classified as mappable and were given an arbitary number of 1, whereas sites not recovered were classified as unmappable with an arbitary number of 0. A 20-kb sliding window was used to determine the running mean mappability, which was

displayed on a scale from 0 (no mappable end per fragment) to 2 (two mappable ends per fragment).

3C-qPCR analysis

The 3C templates of short-term cultured pro-B cells were prepared using HindIII as restriction enzyme as described above for 4C-sequencing (until the first ligation step). DNA (300 ng, quantified by qPCR) of individual 3C templates was subjected to quantitative TaqMan PCR analysis (qPCR) (Hagège et al., 2007). The SensiFAST™ Probe Lo-ROX kit (Bioline Reagents Ltd) was used for qPCR amplification, which was performed with the Bio-Rad CFX Connect™ Real-Time PCR Detection System using region-specific primers and a corresponding LNA® Double-Dye probes with FAM and BHQ1 at the 5' and 3' ends, respectively (listed in Table 2). As internal control for the quality of the 3C template, the ubiquitously expressed *Ercc3* (XPB) locus was analyzed by qPCR (Splinter et al., 2006). The crosslinking frequencies at the *Igh* and control *Ercc3* loci were calculated by using HindIII-digested and randomly ligated BAC DNA of these loci as a standard for PCR amplification. The relative crosslinking frequency was determined as the ratio of the crosslinking frequency at the *Igh* locus relative to the crosslinking frequency at the *Ercc3* gene.

Bio-ChIP-seq analysis of YY1 binding

Pro-B cells isolated from the bone marrow of $Yy1^{ihCd2/+}$ $Rosa26^{BirA/BirA}$ $Rag2^{-/-}$ mice were expanded in vitro for 4-5 days on OP9 cells in the presence of IL-7. About $5x10^7$ pro-B cells were used for chromatin precipitation by streptavidin pulldown (Bio-ChIP), as recently described in detail (Ebert et al., 2011). The precipitated genomic DNA was quantified by real-time PCR, and at least 5 ng of ChIP-precipitated DNA were used as starting material for the generation of paired-end sequencing libraries as described by Illumina's ChIP Sequencing sample preparation protocol followed by size selection of DNA fragments of 200–350 bp. Completed libraries were quantified with the Agilent Bioanalyzer dsDNA 1000 assay kit and Agilent QPCR NGS library quantification kit. Cluster generation and paired-end sequencing was carried out by using the Illumina GAIIx system according to the manufacturer's guidelines, which yielded a read length of 76 bp for each end of a DNA fragment. Peaks were called using the MACS program version 1.3.6.1 (Zhang et al., 2008) with default parameters, a read length of 76, a genome size of 2,654,911,517 bp (mm9) and the appropriate input control sample. Peaks were filtered for p-values of $< 10^{-10}$.

RNA-seq analysis

RNA was isolated from ex vivo sorted pro-B cells using the RNeasy Plus Mini Kit (Qiagen), and mRNA was obtained by two rounds of poly(A) selection using the Dynabeads mRNA Purification Kit (Invitrogen) followed by cDNA preparation and deep sequencing using the Illumina HiSeq 2000 system with a read length of 50 nucleotides, as described (Vilagos et al., 2012). The RNA-seq data were bioinformatically analyzed as described (Vilagos et al., 2012).

3D-FISH analysis

Igh locus-specific DNA probes were prepared from the BACs RP24-376H17 (*Igh* 5' end) and RP24-275L15 (*Igh* 3' end) by nick-translation with dUTP-Alexa568 and dUTP-Alexa488 (Invitrogen), respectively. In vitro short-term cultured pro-B cells were washed in PBS and fixed onto poly-L-lysine-coated slides for 3D DNA-FISH analysis as previously described

(Fuxa et al., 2004; Roldán et al., 2005). Image acquisition was performed by confocal microscopy on a Zeiss LSM780 system with GaAsP detector technology. Optical sections separated by 0.25 µm were collected with a 63x objective (63x/1.4 plan-apochromat Oil DIC). Cells with signals from both alleles were analyzed by the Imaris (Bitplane) software.

Supplemental References

Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. *14*, 1394-1403.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc. Natl. Acad. Sci. USA *100*, 7480-7485.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

Donohoe, M.E., Zhang, L.F., Xu, N., Shi, Y., and Lee, J.T. (2007). Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. Mol. Cell *25*, 43-56.

Driegen, S., Ferreira, R., van Zon, A., Strouboulis, J., Jaegle, M., Grosveld, F., Philipsen, S., and Meijer, D. (2005). A generic tool for biotinylation of tagged proteins in transgenic mice. Transgenic Res. *14*, 477-482.

Dyer, R.B., and Herzog, N.K. (1995). Isolation of intact nuclei for nuclear extract preparation from a fragile B-lymphocyte cell line. BioTechniques *19*, 192-195.

Ebert, A., McManus, S., Tagoh, H., Medvedovic, J., Salvagiotto, G., Novatchkova, M., Tamir, I., Sommer, A., Jaritz, M., and Busslinger, M. (2011). The distal V_H gene cluster of the *Igh* locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells. Immunity *34*, 175-187.

Fuxa, M., Skok, J., Souabni, A., Salvagiotto, G., Roldán, E., and Busslinger, M. (2004). Pax5 induces *V*-to-*DJ* rearrangements and locus contraction of the *immunoglobulin heavy-chain* gene. Genes Dev. *18*, 411-422.

Garrett, F.E., Emelyanov, A.V., Sepulveda, M.A., Flanagan, P., Volpi, S., Li, F., Loukinov, D., Eckhardt, L.A., Lobanenkov, V.V., and Birshtein, B.K. (2005). Chromatin architecture near a potential 3' end of the *Igh* locus involves modular regulation of histone modifications during B-cell development and in vivo occupancy at CTCF sites. Mol. Cell. Biol. *25*, 1511-1525.

Gu, H., Zou, Y.R., and Rajewsky, K. (1993). Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-*loxP*-mediated gene targeting. Cell *73*, 1155-1164.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., *et al.* (2011). CTCF-binding elements mediate control of V(D)J recombination. Nature *477*, 424-430.

Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forné, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat. Protocols *2*, 1722-1733.

Johnston, C.M., Wood, A.L., Bolland, D.J., and Corcoran, A.E. (2006). Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region. J. Immunol. *176*, 4221-4234.

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.

Liu, H., Schmidt-Supprian, M., Shi, Y., Hobeika, E., Barteneva, N., Jumaa, H., Pelanda, R., Reth, M., Skok, J., Rajewsky, K., *et al.* (2007). Yin Yang 1 is a critical regulator of B-cell development. Genes Dev. *21*, 1179-1189.

McCormack, M.P., Forster, A., Drynan, L., Pannell, R., and Rabbitts, T.H. (2003). The LMO2 T-cell oncogene is activated via chromosomal translocations or retroviral insertion during gene therapy but has no mandatory role in normal T-cell development. Mol. Cell. Biol. *23*, 9003-9013.

McManus, S., Ebert, A., Salvagiotto, G., Medvedovic, J., Sun, Q., Tamir, I., Jaritz, M., Tagoh, H., and Busslinger, M. (2011). The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. EMBO J. *30*, 2388-2404.

Nutt, S.L., Urbánek, P., Rolink, A., and Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced *V*-to-*DJ* recombination at the *IgH* locus. Genes Dev. *11*, 476-491.

Perlot, T., Alt, F.W., Bassing, C.H., Suh, H., and Pinaud, E. (2005). Elucidation of IgH intronic enhancer functions via germ-line deletion. Proc. Natl. Acad. Sci. USA *102*, 14362-14367.

Retter, I., Chevillard, C., Scharfe, M., Conrad, A., Hafner, M., Im, T.H., Ludewig, M., Nordsiek, G., Severitt, S., Thies, S., *et al.* (2007). Sequence and characterization of the Ig heavy chain constant and partial variable region of the mouse strain 129S1. J. Immunol. *179*, 2419-2427.

Roldán, E., Fuxa, M., Chong, W., Martinez, D., Novatchkova, M., Busslinger, M., and Skok, J.A. (2005). Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. Nat. Immunol. *6*, 31-41.

Shinkai, Y., Rathbun, G., Lam, K.-P., Oltz, E.M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., Stall, A.M., *et al.* (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell *68*, 855-867.

Simonis, M., Kooren, J., and de Laat, W. (2007). An evaluation of 3C-based methods to capture DNA interactions. Nature Methods *4*, 895-901.

Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. Genes Dev. *20*, 2349-2354.

Tallquist, M.D., and Soriano, P. (2000). Epiblast-restricted Cre expression in MORE mice: a tool to distinguish embryonic vs. extra-embryonic gene function. Genesis *26*, 113-115.

Urbánek, P., Wang, Z.-Q., Fetka, I., Wagner, E.F., and Busslinger, M. (1994). Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. Cell *79*, 901-912.

van de Werken, H.J., de Vree, P.J., Splinter, E., Holwerda, S.J., Klous, P., de Wit, E., and de Laat, W. (2012). 4C technology: protocols and data analysis. Meth. Enzymol. *513*, 89-112.

Vilagos, B., Hoffmann, M., Souabni, A., Sun, Q., Werner, B., Medvedovic, J., Bilic, I., Minnich, M., Axelsson, E., Jaritz, M., *et al.* (2012). Essential role of EBF1 in the generation and function of distinct mature B cell types. J. Exp. Med. *209*, 775-792.

Vincent-Fabert, C., Fiancette, R., Pinaud, E., Truffinet, V., Cogné, N., Cogné, M., and Denizot, Y. (2010). Genomic deletion of the whole IgH 3' regulatory region (hs3a, hs1,2, hs3b, and hs4) dramatically affects class switch recombination and Ig secretion to all isotypes. Blood *116*, 1895-1898.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.