Supporting Information

Verma-Gaur et al. 10.1073/pnas.1208398109

SI Materials and Methods

Mice and Cell Lines. *Rag1*$^{-/-}$, *YY1*$^{f/f}$ x *mb1-Cre*, and *Rag1*$^{-/-}$*YY1*$^{f/f}$ x *mb1-Cre* mice, all on a C57BL/6 background (1), and *Rag1*$^{-/-}$ mice on the C.129 background were maintained at the breeding colony at The Scripps Research Institute (TSRI) in accordance with protocols approved by the TSRI Institutional Animal Care and Use Committee. Breeding pairs of *YY1*$^{f/f}$ × *mb1-Cre* mice were kindly provided by H. Liu and Y. Shi (Harvard University, Cambridge, MA) (2). Eμ-deleted mice were kindly provided by F. Alt and C. Guo (Harvard University, Cambridge, MA) (3). The C57BL/6 murine embryonic fibroblast (MEF) cells were obtained from K. Mowen (TSRI, La Jolla, CA). *E2A*$^{-/-}$ prepro-B cells were cultured as previously described (4). R2K are Abelson-MuLV derived cells lines from *Rag2*$^{-/-}$ mice on C57BL/6 background, and were kindly provided by C. Bassing (University of Pennsylvania, Philadelphia, PA).

Production and Transduction of CTCF shRNA retroviruses. Retroviral plasmids containing CTCF shRNA target and control sequences were generously provided by C. Wilson (University of Washington, Seattle, WA) (5). B6 pro-B cells were transduced with viral supernatant using spin infection on day 4 of IL-7 culture, as previously described (4). The GFP^{+} CD19^{+} pro-B cells were isolated by cell sorting 4 d after retroviral transduction. Knockdown of CTCF protein and RNA was confirmed by Western blot and by RT-PCR, as previously described (4).

RT-PCR Analysis. Total RNA was isolated from from freshly isolated pro-B cells from *Rag1*$^{-/-}$, *YY1*$^{-/-}$ *Rag1*$^{-/-}$, and Eμ$^{-/-}$ mice by the TRIzol method (Roche). RNA was reverse-transcribed using QuantiTect Reverse Transcription kit (Qiagen) and random hexamer primers (Roche). The relative transcription levels were measured by real-time PCR of cDNA samples and were normalized using β-actin RNA control. For J558 sense-specific PCR, 300 ng of total RNA was reverse-transcribed using a J558 sense-specific primer (GAGCTTGCTGCACCTCCA) followed by real-time PCR. The list and sequences of primers used for real-time PCR are provided in Table S1.

ChIP and ChIP-Sequencing. ChIP and ChIP-seq samples were prepared from freshly isolated *Rag1*$^{-/-}$ pro-B cells, as previously described (4). The following antibodies were used for ChIP: anti-YY1 (Abcam; ab12132 and Santa Cruz; H-414), anti-Pax5 (Santa Cruz; C-20), anti-Pol II CTD (phospho S2) (Abcam; ab5095) and H3K4me3 (Active Motif). The relative binding of YY1 and Pax5 were measured by real-time quantitative PCR (qPCR) of ChIP samples and were normalized using actin control. The list of primers is shown in Table S2. For the ChIP-seq, input and immunoprecipitated DNA was given to the Scripps DNA Array Facility, where it was prepared for massively parallel sequencing on Illumina Genome Analyzer IIx. The sequencing libraries were prepared using 10 ng DNA and using protocol as previously described (4).

ChIP-Seq Analysis. The Bowtie algorithm was used to map experimental and input control fastaq tags to the mm9 genome. Duplicated reads were removed before mapping. All valid alignments were reported using the most exhaustive search settings with a maximum of two mismatches allowed within the seed region of each tag (6). Tags mapping uniquely to the genome, defined as reads where the best alignment score occurs at only a single locus in the mm9 genome, were extended 300-bp upstream or downstream from the alignment start site, depending upon the strand orientation of the mapping. All reads mapping with their best alignment score at multiple sites in the genome were discarded unless all of those sites resided within LINE elements, Ig regions, or T-Cell receptor regions.

To identify ChIP-seq peaks, the mm9 genome was divided into nonoverlapping 25-bp bins, and the number of experimental and input control tags mapping to each bin were counted. Peak probabilities at each bin were calculated using the Poisson distribution where a dynamic λ-parameter is chosen to correct for potential biases in the background distribution of control tags stemming from sequencing bias, mapping bias, chromatin structure, and other sources of bias as described in the MACS algorithm (7). The dynamic λ-parameter was chosen at each bin from the largest of the total average of control tags per bin, or the average control tags per bin in a 1-, 5-, or 10-kb window surrounding each bin. This parameter was finally corrected for the difference in the number of experimental vs. control tags to calculate a final peak probability. Peaks are formally defined as bins with an enrichment of experimental tags at a probability less than 1×10^{-5} according to the Poisson probability.

Directional RNA-Sequencing and Library Preparation. Ten micrograms of total RNA from freshly isolated *Rag1*$^{-/-}$ pro-B cells was depleted of ribosomal RNA (rRNA) using RiboMinus (Invitrogen) kit. The depletion was verified using Agilent 2100 Bioanalyser. Approximately 200 ng of Ribo-minus–treated total RNA was prepared for Next Generation Sequencing using a slightly modified version of the Illumina protocol (Directional_mRNA-Seq_SamplePrep_Guide_15018460_A.pdf), in which 12 cycles of PCR were performed instead of 15 and standard Truseq adapters and Truseq barcoded primers were used instead of the v1.5 adapters and primers described in the guide. A final size selection was performed using agarose gel to capture a library with insert sizes ranging from 50–150 bases in length suitable for 40-base single-read sequencing. The library product was purified from agarose using standard oligo purification columns. The prepared library was then loaded onto an Illumina HiSeq v1.5 single-read flowcell, standard-cluster generation was performed on a Cbot and sequenced for 40 bases of the insert and 7 bases of the index read using standard HiSeq sequencing reagents. After sequencing, reads were processed using CASAVA 1.8 and demultiplexed based on index sequences.

For preparation of libraries for target selection on Agilent SureSelect DNA Capture Arrays, we followed the procedure described above except that we selected insert sizes to be 150–200 bases in length. In addition, after the first round of gel purification, eight additional cycles of PCR was performed, followed by another gel-purification step, then 12 more cycles of PCR followed by Ampure XP bead purification. cDNA was hybridized to arrays using Agilent protocols. After elution of target from the arrays, the sample was PCR-amplified for 18 additional cycles (based on Agilent's protocols). The library was gel purified before loading onto the flow-cell for sequencing the 100-bp reads.

RNA-Seq Analysis. The initial data analysis of 40-bp and 100-bp RNA-seq was performed using the Genome Analyzer Pipeline Software (Casava 1.8.1) including the image analysis, base calling, and alignment. The reads were adaptor-trimmed using Flicker 3.0, which is an Illumina proprietary early access software. TopHat (http://tophat.cbcb.umd.edu), a splice junction mapper for RNA-Seq reads, was used for read alignment to the Mouse genome

Build version mm9. BEDtools v2.13.3 was used to convert the output BAM file to BED12 format. Reads to Chromosome 12 were extracted and partitioned by plus- and minus-strand reads. The junction reads were split and any segments >25 bp were retained. The mapped reads were then sorted and the duplicates with the same start and end positions were removed. Finally, the BED files were converted to BEDgraph files using the BEDtools function genomeCoverageBed.

Heat-Shock and DRB Treatment of Pro-B Cells. Freshly isolated *Rag1*$^{-/-}$ pro-B cells were subjected to heat shock at 45 °C for 30 min, or were incubated with 100 μM dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) for 3 h. Cells were immediately harvested for RNA or for chromosome conformation capture (3C) lysate. Cells were greater than 90% viable at the time of harvest.

The 3C Analysis. The 3C analysis was performed as previously outlined using HindIII to digest cross-linked chromatin (4). Digestion efficiency was calculated as described in Hagège et al. using primers described previously (8). Only samples with HindIII efficiencies greater than 80% were included in the 3C assays. We used a control template containing all possible ligation products to correct for differences in PCR efficiency between different templates. The control template contained equimolar amounts of the BAC clones spanning the Pax5-activated intergenic repeat (PAIR) 4 (RP23-101G12), PAIR6 (RP23-354D10), 3′RR (RP23-149L24), and D_H (RP23-270B12) portions of the *Igh* locus and PCR amplification templates of the Calreticulin (CalR) locus. To compare results from different cell types or cells treated with different retroviral shRNAs, the results were normalized to the ligation frequency of two restriction fragments in the CalR locus. Primers are provided in Table S3. Statistical analysis is shown in Table S4.

1. Hobeika E, et al. (2006) Testing gene function early in the B cell lineage in mb1-cre mice. *Proc Natl Acad Sci USA* 103:13789–13794.
2. Liu H, et al. (2007) Yin Yang 1 is a critical regulator of B-cell development. *Genes Dev* 21:1179–1189.
3. Perlot T, Alt FW, Bassing CH, Suh H, Pinaud E (2005) Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc Natl Acad Sci USA* 102:14362–14367.
4. Degner SC, et al. (2011) CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc Natl Acad Sci USA* 108:9566–9571.
5. Sekimata M, et al. (2009) CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. *Immunity* 31:551–564.
6. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.
7. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137.
8. Hagège H, et al. (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2:1722–1733.

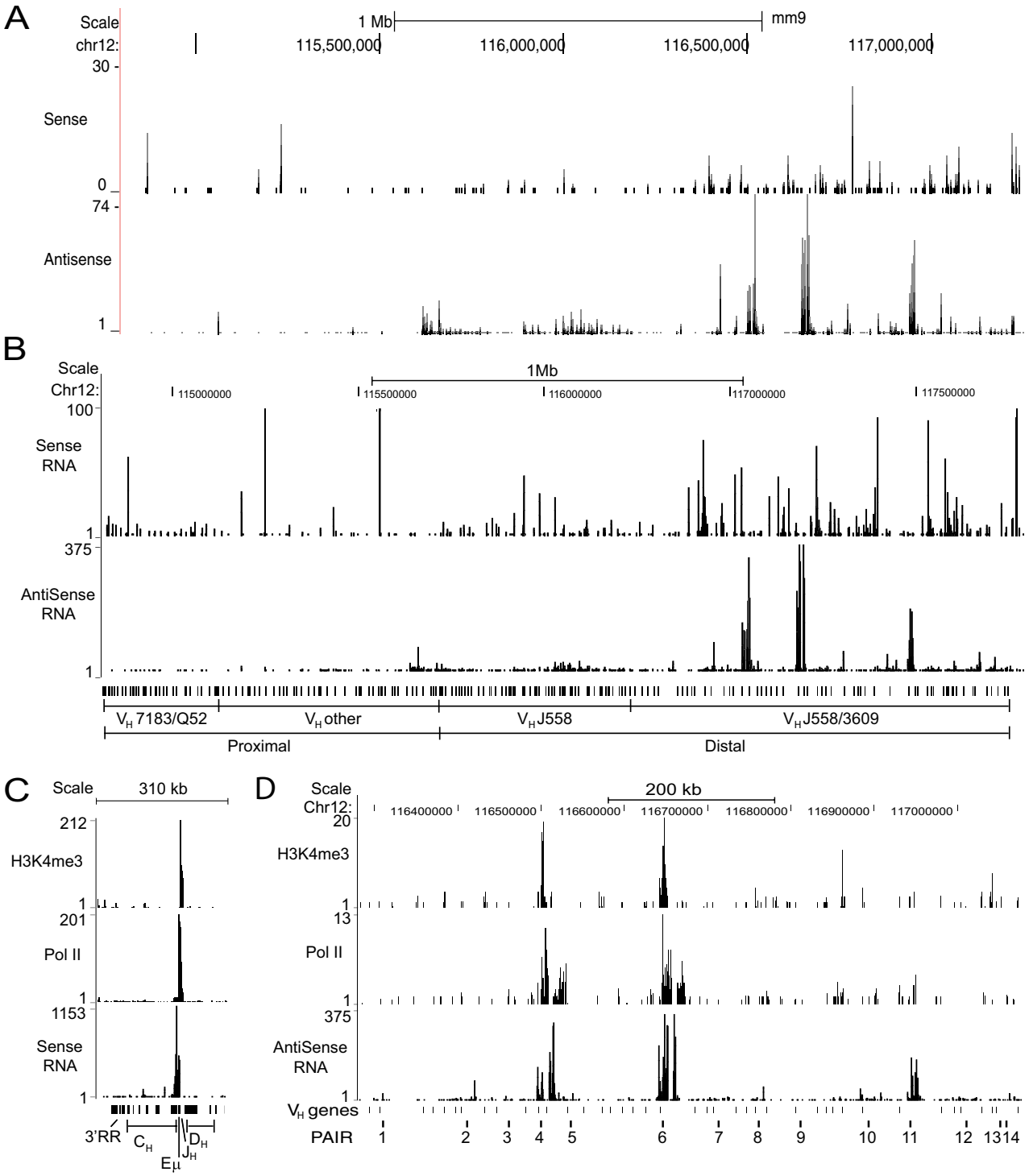

Fig. S1. RNA-seq analysis of *Rag1*$^{-/-}$ pro–B-cell RNA. (*A*) Levels of sense and antisense transcripts throughout the V_H region of *Igh* locus are shown from the first RNA-seq. (*B*) Levels of sense and antisense transcripts throughout the V_H region of *Igh* locus from the RNA-seq performed with the array-enriched RNA. (*C*) ChIP-seqs for H3K4me3, Pol II, and sense RNA in the Eμ region of the *Igh* locus. (*D*) ChIP-seqs for H3K4me3, Pol II, and sense RNA in the PAIR regions.

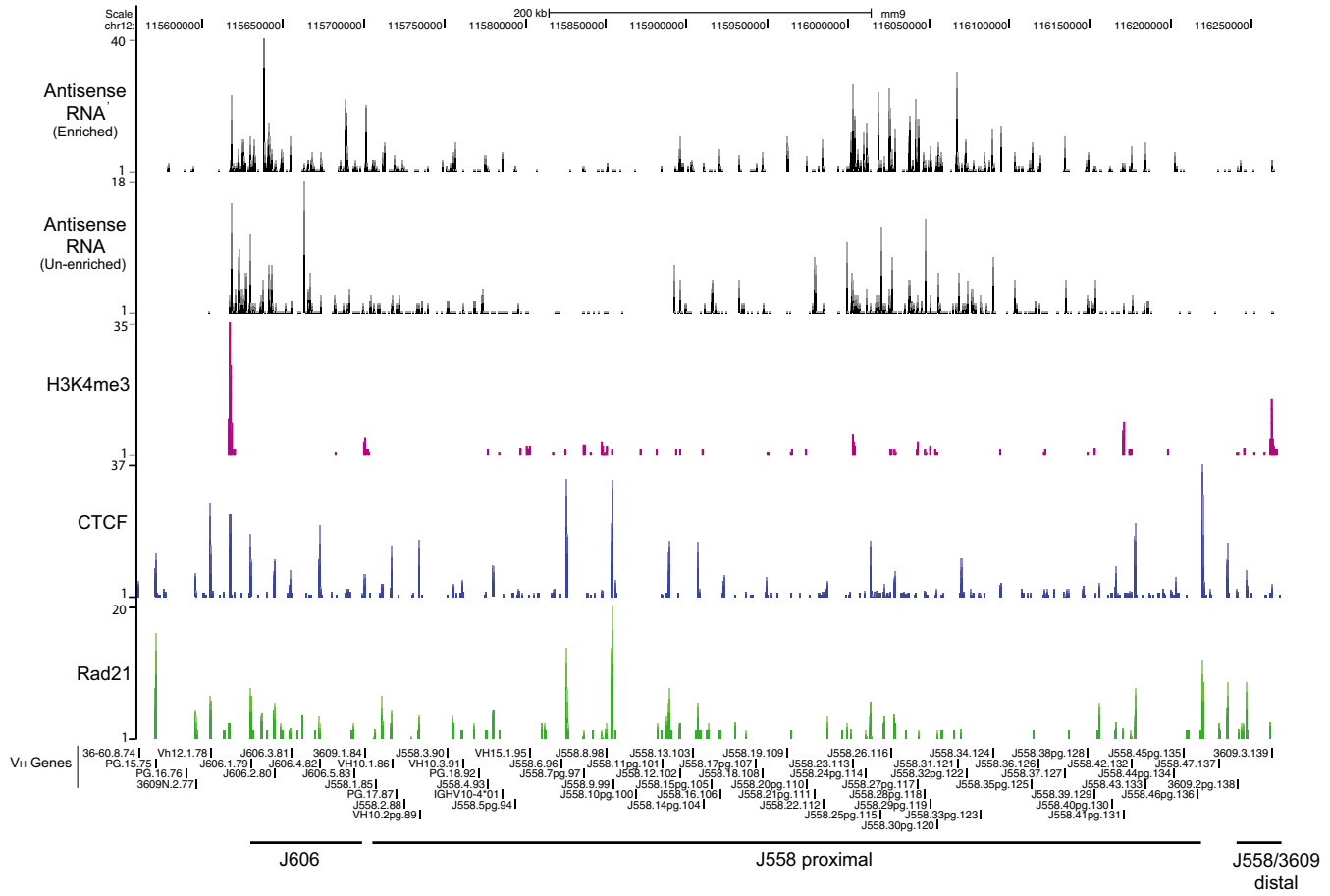

Fig. S2. RNA-seq analysis of the middle of the V_H locus showing other regions of antisense transcription. The two more minor regions of antisense RNA, in the proximal J558 region and in the J606 region, can be seen. Both the original RNA-seq and the Agilent array-enriched RNA seq are displayed. ChIP-seq data for H3K4me3 shows high level at the start of the J606 antisense. Interestingly, there are high peaks of cohesin/CTCF flanking the J606 and J558 antisense transcription, but the level of CTCF/cohesin is lower over the highly transcribed regions. Thus, the CTCF/cohesin complex may form domains or boundaries that regulate other antisense transcription.

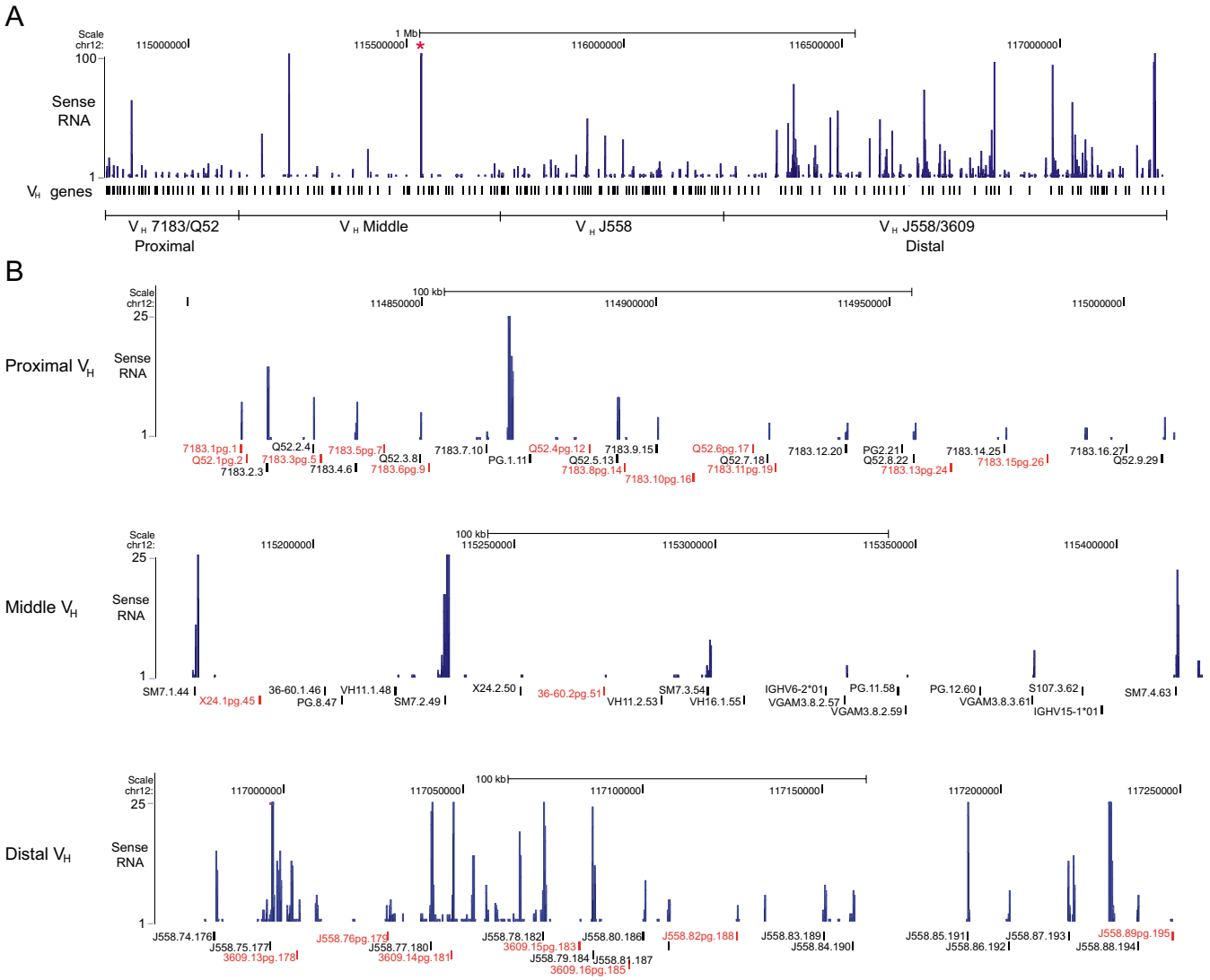

Fig. S3. Sense transcripts are expressed throughout the V_H locus. (*A*) Levels of sense transcripts throughout the V_H region of *Igh* locus are shown from the RNA-seq performed with the array-enriched RNA. The red asterisk shows the peak that was absent in the nonenriched RNA-seq and was not expressed when tested by RT-PCR. (*B*) Close-up pictures of different regions of *Igh* locus showing sense transcripts from the proximal, middle, and distal V_H region. The scale is set to 25 to help visualize the lower level peaks in the proximal and middle regions, but which will truncate the taller peaks. Pseudogenes are shown in red. It can be seen that transcripts are rare from V_H genes in the middle portion of the V_H locus other than relatively high levels of reads at SM7 genes. The peaks are the highest in the J558 portion of the locus, and occasionally multiple sets of peaks are present around the coding region.

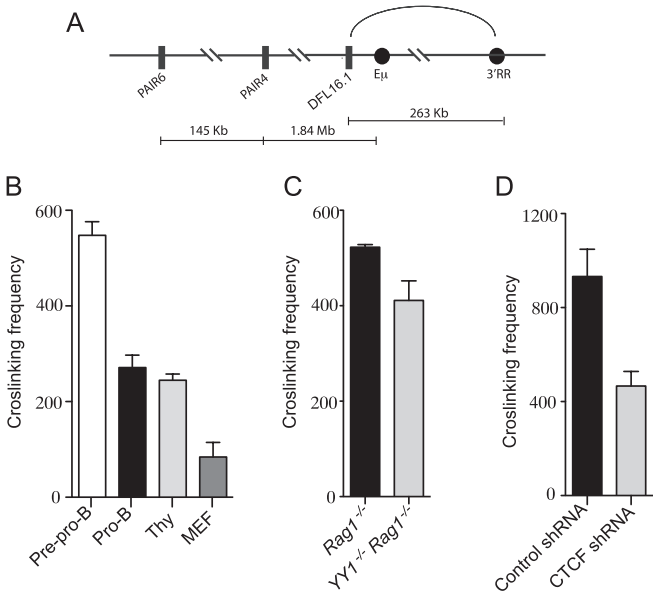

Fig. S4. The 3C analysis showing the interaction between CTCF/DFL and CTCF/3′RR. (*A*) Schematic diagram showing the position of CTCF/DFL and CTCF/3′RR region in *Igh* locus. (*B*) Cross-linking frequency between CTCF/DFL and CTCF/3′RR was measured using the CTCF/DFL Taqman probe in *E2A*$^{-/-}$ prepro-B cells, *Rag1*$^{-/-}$ pro-B cells, thymocytes and MEFs. Data are presented as mean ± SEM, *n* = 2–3. (*C*) Cross-linking frequency between CTCF/DFL and CTCF/3′RR in *YY1*$^{-/-}$*Rag1*$^{-/-}$ pro-B cells compared with *Rag1*$^{-/-}$ pro-B cells. Data are presented as mean ± SEM, *n* = 3. (*D*) Cross-linking frequency between CTCF/DFL and CTCF/3′RR in the A-MuLV transformed *Rag2*$^{-/-}$ cell line R2K transduced with control (scramble) or CTCF targeting shRNA retroviruses. Data are presented as mean ± SEM, *n* = 4.

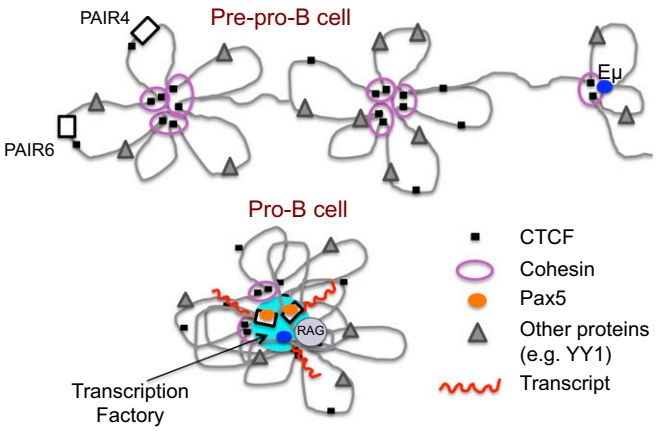

Fig. S5. Model of 3D changes in the architecture of the *Igh* locus. In prepro-B cells, the *Igh* locus has been proposed to be in a multirosette structure. CTCF is likely at the base of the loops of the rosettes. Pax5 is not bound to PAIR elements, and there is no germ-line transcription. In the pro–B-cell stage, Pax5 binds to the PAIR elements, PAIR transcripts are made from PAIR4 and PAIR6, as well as from other antisense promoters, and the sense promoters of many V_H genes. Eμ is constantly being transcribed because of sense germ-line transcription, and thus is likely always in a transcription factory, and D-J_H is within 2 kb of Eμ. We propose that the creation of transcription-mediated long-range interactions at transcription factories is a key role of noncoding antisense and sense transcription, and that the antisense transcription might therefore initially alter the 3D conformation of the *Igh* locus, bringing the distal part of the locus close to Eμ. This process could be followed by additional stochastic and dynamic interactions from the many sense and antisense germ-line promoters throughout the *Igh* locus moving in and out of transcription factories, further compacting the locus, and bringing V_H genes throughout the locus into close proximity to the DJ_H region to which one V_H gene will rearrange. Our previous studies have shown that CTCF plays an architectural role in the general 3D structure of the *Igh* locus and acted as negative regulator of antisense transcription from distal V_HJ558 intergenic region. We propose that CTCF is critical for the formation of the multilooped rosette-like structure that is present in prepro-B cells as well as in pro-B cells. We suggest that this basic architectural role of the CTCF/cohesin complex is distinct from the role that it may play in the 3D structure of the *Igh* locus specifically in pro-B cells. We have previously shown that there is increased cohesin binding to CTCF sites within the V_H locus in pro-B cells. These CTCF/cohesin complexes may act in concert with transcription factors such as Pax5 and YY1 to produce full locus compaction, in part by regulating the extent of noncoding transcription. The CTCF sites at PAIR elements are in between the start site of transcription and the Pax5 and E2A binding sites in the promoter. Given that knockdown of CTCF increases antisense transcription, we propose that the CTCF/cohesin complexes act as insulators there, and regulate the level of PAIR antisense transcription. This insulator role would be in addition to the architectural role of CTCF/cohesin complexes.

Table S1. Primers used for RT-PCR

Name	Primer sequence (5′ to 3′)
PAIR4 (1)	F: TCCATGTTAGTGGTGGCAGA
	R: GTGACGACGGCTCATGACTA
PAIR6 (1)	F: TCCATGTTAGTGGTGGCAGA
	R: TCTGCAGTGTGTGACGACAG
J558 5′Int (2)	F: ATTCCCCTCCCAATAGGAAA
	R: TGTCAATCACAATGGGCATC
J558 3′Int (2)	F: GCCAATAGGAAAGCAGGTGA
	R: TGGAAAAGTTGCAGTCAATCA
J558 sense (2)	F: ATGGGATGGAGCTGGATCTT
	R: GACACACTCAGGATGTGTTTGTAG
J606	F: CACACATTATCTGTGAGTAGAGATT
	R: GGGTGACATCTCTGACTACTCT
μ^0 (3)	F: TTAACCGAGGAATGGGAGTG
	R: GGTGGGGCTGGACAGAGTGT
HSP70	F: CAGACTCTTTGCACTTGATAGCTG
	R: CACAGTGCTGCTCCCAACATTAC
Pax5	F: GTCCCAGCTTCCAGTCACAG
	R: AATAGGGTAGGACTGTGGGC
18s RNA (4)	F: TTGACGGAAGGGCACCACCAG
	R: GCACCACCACCCACGGAATCG
β-Actin	F: TGTTACCAACTGGGACGACA
	R: GGGGTGTTGAAGGTCTCAAA

1. Ebert A, et al. (2011) The distal VH gene cluster of the Igh locus contains distinct regulatory elements with pax5 transcription factor-dependent activity in pro-B cells. *Immunity* 34(2): 175–187.
2. Bolland DJ, et al. (2004) Antisense intergenic transcription in V(D)J recombination. *Nat Immunol* 5(6):630–637.
3. Bolland DJ, et al. (2007) Antisense intergenic transcription precedes Igh D-to-J recombination and is controlled by the intronic enhancer Emu. *Mol Cell Biol* 27(15):5523–5533.
4. Rhinn H, et al. (2008) Housekeeping while brain's storming Validation of normalizing factors for gene expression studies in a murine model of traumatic brain injury. *BMC Mol Biol* 9: 62.

Table S2. Primers used for ChIP assay

Name	Closest V_H gene	Primer sequence (5′ to 3′)
hs7 (1)		F: TCTGCGCCTGCGTAAGAGGA
		R: GAACAGGAACTGCGACTGGAG
hs6 (1)		F: AGCGTCTGCGTGGTGTGTG
		R: TACATCTGCCTCAGCAGCTCAG
Eμ (1)		F: GCTGCAGGTGTTCCGGTTCTGATCGGCC
		R: GGAATGGGAGTGAGGCTCTCT
CTCF/DFL (1)		F: GAGGGTTGTCACAGTTGACAC
		R: CATCTTTCTGAAGCTCACACTAC
V_H1 (1)	7183.4.6	F: GGCTGTGTCCTCAGTCTGCA
		R: CTAAAACCTCCTGCAGAGCAC
V_H3 (1)	7183.14.25	F: TCTAATTGTACAGGCTGTAACTC
		R: CTCAAACTAAAACCTGCAGAGC
V_HA	SM7.4.63	F: CACGCCAGTGAAGACGTTCACTAC
		R: GCTGCCTGGAACTGAGATGACAG
V_HB	J606.1.79	F: CACAACATGAACCAGTGCCAGTC
		R: CCGAGCGCACCCTAAGAGAAC
V_HC	J558.8.98	F: GAGTGTGACTGCAGGATCACCAG
		R: CACCAAAGCACAAGAGCTTCCC
V_HD	3609.3.139	F: CACACCACCCATATAGACAGCTGCAT
		R: GTTTGGGCTGGTTCTCAGCTCTG
J558 5′Int (2)		F: ATTCCCCTCCCAATAGGAAA
		R: TGTCAATCACAATGGGCATC
J558 3′Int (2)		F: GCCAATAGGAAAGCAGGTGA
		R: TGGAAAAGTTGCAGTCAATCA
PAIR4/6/11 (3)		F: AGAATGTCAGCTTACCCATTCAA
		R: AATCCTGTCACAGGAGTTCCA
PAIR7/12/2 (3)		F: CATTCAAGAACTGAAGGAGAGC
		R: TGGTAGTCTTGCCACAGGAG
V_H10 (1)	J558.69.170	F: CAGCATCTCTCTGCTGACCA
		R: GAACCAAGTAAGCACACTCTTCAG
γ-Actin		F: AGGCATGGAGTCCTGTGGTATC
		R: AGCCACAGGTCCTAAGGCCAG
CD19		F: GATTTGGAAGAGTGCCTGCA
		R: GCCTGCCTCCTACTAAGGTA
BLNK (3)		F: ACACAACGCACAGTAATCAGC
		R: CCTGGACTTCCTGGGGAGCAG
HSP70		F: CAGACTCTTTGCACTTGATAGCTG
		R: CACAGTGCTGCTCCCAACATTAC

1. Degner SC, Wong TP, Jankevicius G, Feeney AJ (2009) Cutting edge: Developmental stage-specific recruitment of cohesin to CTCF sites throughout Ig loci during B lymphocyte development. *J Immunol* 182(1):44–48.
2. Bolland DJ, et al. (2004) Antisense intergenic transcription in V(D)J recombination. *Nat Immunol* 5(6):630–637.
3. Ebert A, et al. (2011) The distal VH gene cluster of the Igh locus contains distinct regulatory elements with pax5 transcription factor-dependent activity in pro-B cells. *Immunity* 34(2): 175–187.

Table S3. Primers used for 3C assay

Name	Primer sequence 5′ to 3′	Mm9 location
Ligation control primers		
CalR fragB (1) F	CCCAAACCACCACTACCATTACA	chr8: 87363228–87363250
CalR fragA F	GATGAACTGCCCTATCCTGAGTC	chr8: 88183257–88183279
TaqMan probes used for 3C assay		
CTCF/DFL (2)	TCTCAGTCTAGGGCTGCTGAGTGATTCC	chr12: 114723017–114723044
Eμ*	TGGCTTACCATTTGCGGTGCCTGGTTT	chr12: 114664858–114664884
3C interaction primers		
A	TTCTGAGGACCAGGAAGGAACCA	chr12: 117037330–117037352[†]
PAIR6	CATTTCTGAATGTAGTCGAATTATACTC	chr12: 116643843–116643870
B	GCCAGATCTCAAGCCTGCCACACCTC	chr12: 116610548–116610573
PAIR4	CATGCCAAAAACACTACATAAAGTTG	chr12: 116498042–116498068
C	GATCACTATACTTGCTGGTCTGGTGCAATG	chr12: 116490467–116490496 chr12: 116633996–116634025
Eμ [T2(2P)]*	TCCACACAAAGACTCTGGACCTCT	chr12:114664888–114664911
CTCF/3′RR (SD120) (2)	CTCCACAATGACCACAGCGT	chr12:114459642–114459661

*Primer and probe sequences were kindly provided by Amy Kenter (University of Illinois, Chicago, IL)
[†]Primer A also binds to two other sites which are ~3–8 kb away from the HindIII restriction site. Therefore, it cannot be amplified in TaqMan PCR.

1. Ju Z, et al. (2007) Evidence for physical interaction between the Ig heavy chain variable region and the 3′ regulatory region. *J Biol Chem* 282:35169–35178.
2. Degner SC, et al. (2011) CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc Natl Acad Sci USA* 108:9566–9571.

Table S4. Statistics

Figure comparison	For:	*P* value
Fig. 2*A*		
Rag$^{-/-}$ vs. *YY1*$^{-/-}$ *RAG*$^{-/-}$	PAIR4	0.0097
	PAIR6	0.0097
	J558 5′Int	0.0473
	J558 3′Int	0.0462
	J558 Sense	0.1932
	J606	0.036
Fig. 3*B*		
Pro-B vs. prepro-B	PAIR6	0.0009
Pro-B vs. Thy	PAIR6	0.031
Pro-B vs. MEF	PAIR6	0.0418
Pro-B vs. prepro-B	PAIR4	0.01
Pro-B vs. Thy	PAIR4	0.0325
Pro-B vs. MEF	PAIR4	0.0027
Fig. 3*C*		
Rag$^{-/-}$ vs. *YY1*$^{-/-}$*Rag*$^{-/-}$	PAIR6	0.0235
Rag$^{-/-}$ vs. *YY1*$^{-/-}$*Rag*$^{-/-}$	PAIR4	0.0832
Fig. 3*D*		
CTCF vs. control shRNA	PAIR6	0.1139
CTCF vs. control shRNA	PAIR4	0.1504
Fig. 3*E*		
RAG$^{-/-}$ vs. Eμ	J558 5′Int	0.28
RAG$^{-/-}$ vs. Eμ	PAIR4	0.235

All were analyzed by a two-tailed *t* test.