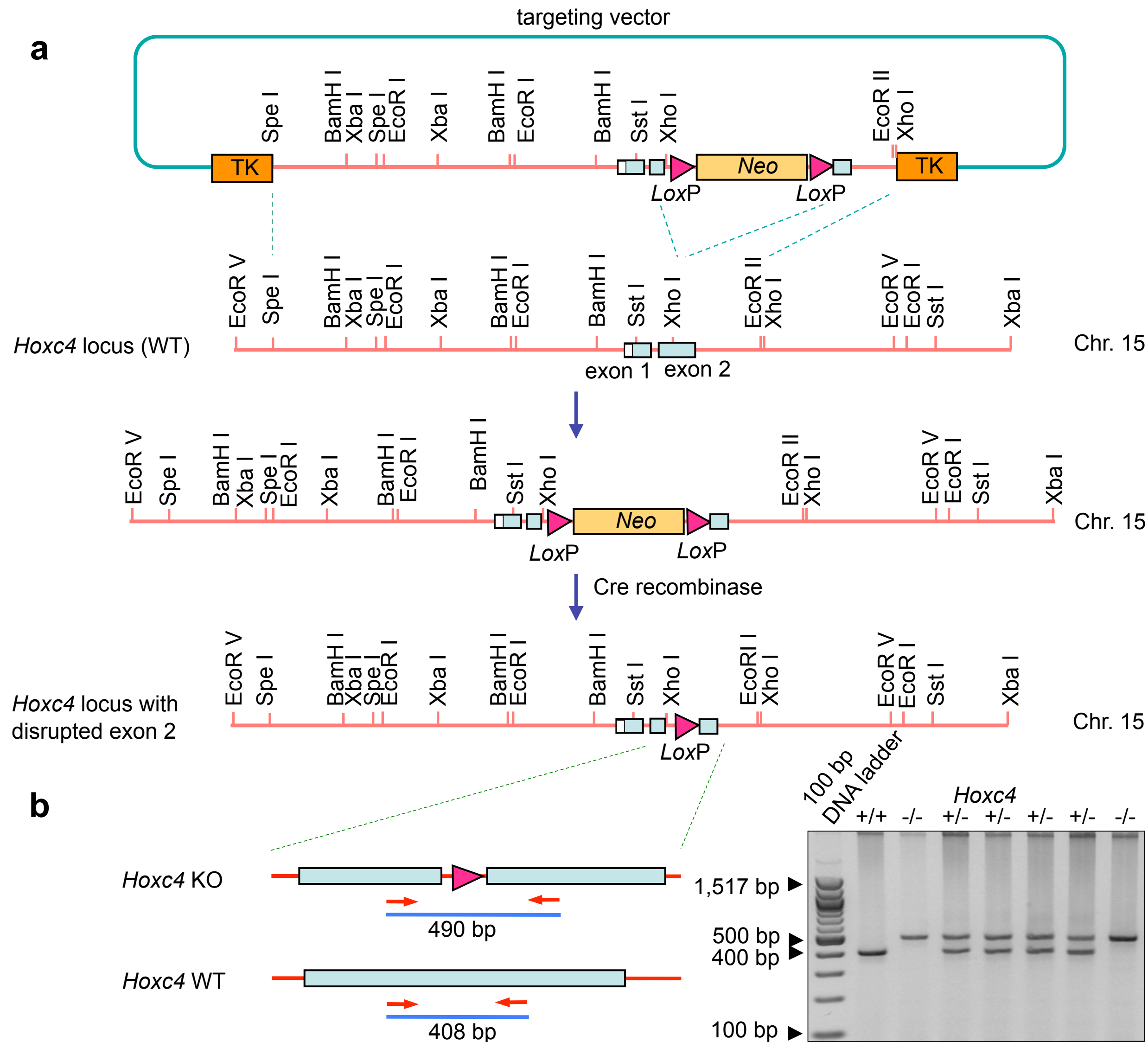
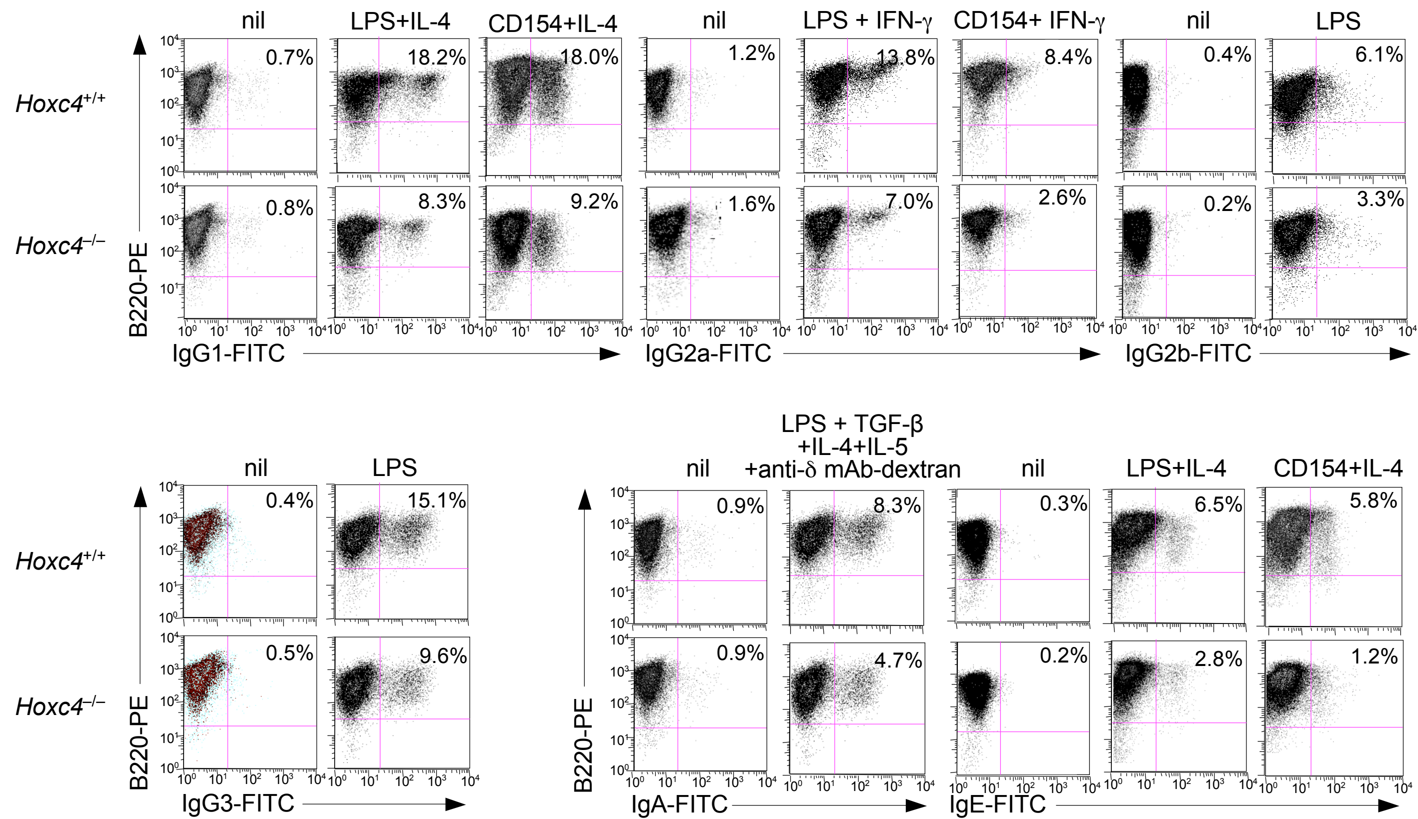
HoxC4 binds to the *AICDA/Aicda* promoter to induce AID expression, class switch DNA recombination and somatic hypermutation

Seok-Rae Park, Hong Zan, Zsuzsanna Pal, Jinsong Zhang, Ahmed Al-Qahtani, Egest J Pone, Zhenming Xu, Thach Mai & Paolo Casali

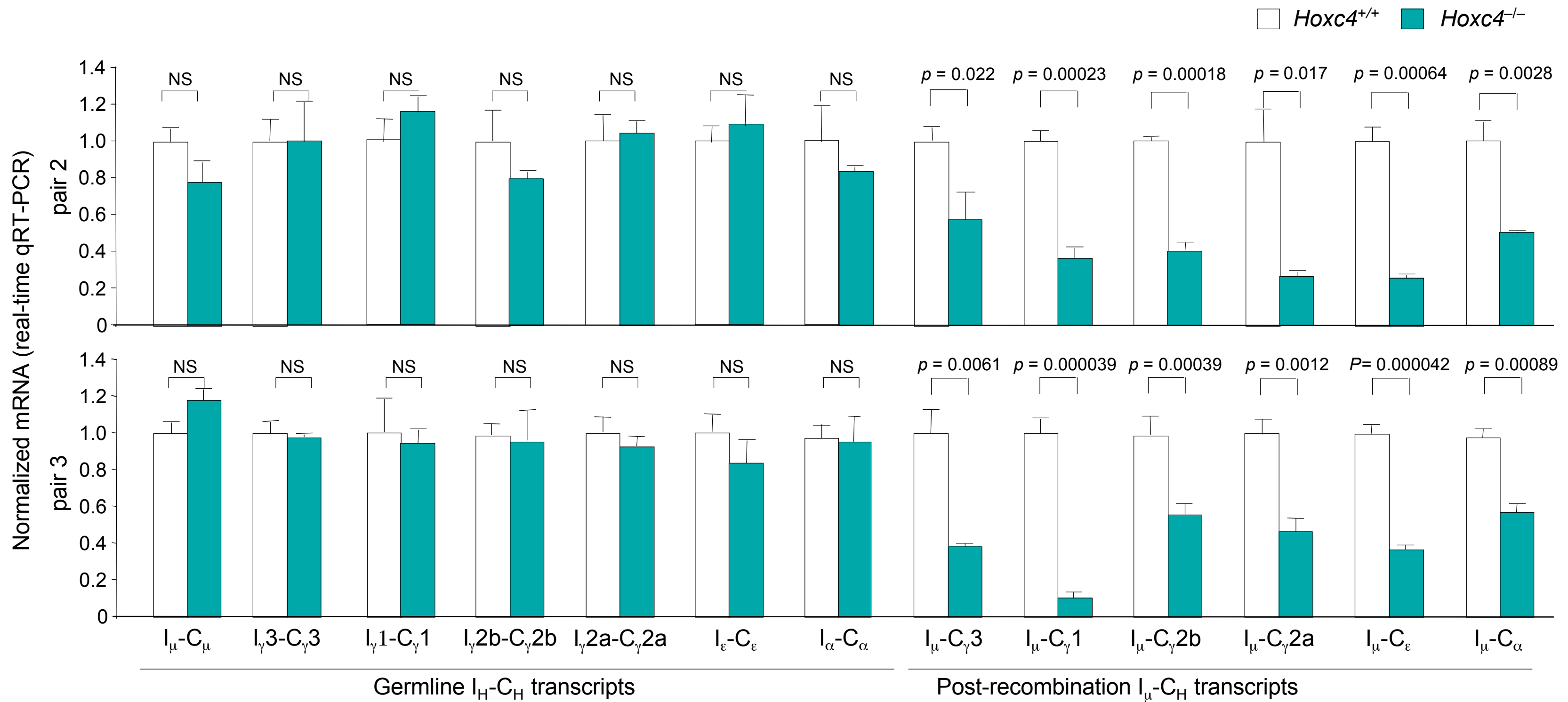
Institute for Immunology, School of Medicine and School of Biological Sciences, University of California, 3028 Hewitt Hall, Irvine, CA 92697-4120, USA.



Supplementary Figure 1. (**a**) Construction of the *Hoxc4* knockout locus (adapted from A. M. Boulet and M. R. Capecchi, unpublished). A *lox*P-flanked *neo*^{r} cassette was inserted in exon 2 of the *Hoxc4* gene at the *Xho* **I** site between the third and the fourth codons of the *Hoxc4* homeobox. This is the same insertion site as in the published *Hoxc4* knockout locus (Boulet and Capecchi. 1996. *Dev. Biol*. 177:232-249). Subsequently, the *neo*^{r} cassette was removed by Cre recombination, so that only a single *lox*P site remains. This introduces a stop codon at the insertion site, yielding a nonfunctional truncated protein (lacking 95% of the homeodomain). (**b**) PCR analysis for the WT and genetically disrupted (KO) *Hoxc4* locus of the progeny from a cross between two *Hoxc4*$^{+/-}$ mice, using the forward and reverse *Hoxc4* primers (small red arrows) upstream and downstream of the insertion site, respectively. PCR of the WT *Hoxc4* locus generated a 408 bp product, PCR of the disrupted *Hoxc4* locus generated a bigger (490 bp) product, as it included the inserted *lox*P site. PCR products of WT and KO *Hoxc4* locus were verified by sequencing.



Supplementary Figure 2. Impaired CSR in *Hoxc4*$^{-/-}$ B cells. Spleen *Hoxc4*$^{+/+}$ and *Hoxc4*$^{-/-}$ B cells were stimulated with LPS or CD154 in the presence of nil (for IgG2b and IgG3), IL-4 (for IgG1 and IgE), IFN-γ (for IgG2a), or TGF-β1, IL-4, IL-5 and anti-δ mAb-dextran (for IgA). After a 4-d culture, the cells were stained with PE-anti-mouse B220 mAb and FITC-anti-mouse IgG1, IgG2a, IgG2b, IgG3, IgA or IgE mAb for surface analysis (the number inside each panel indicates the percentage of B220^{+} cells that are positive for the indicated Ig isotypes).



Supplementary Figure 3. HoxC4 deficiency does not alter the level of germline I_H-C_H transcripts but reduces post-recombination Iμ-C_H transcripts. Spleen *Hoxc4*$^{+/+}$ and *Hoxc4*$^{–/–}$ B cells were cultured with LPS or LPS and cytokines for 3 d and then harvested for RNA extraction. This was used as template in real-time qRT-PCR to measure the levels of germline I_μ-C_μ, I_γ3-C_γ3, $I_{\gamma 1}$-C_γ1, I_γ2b-C_γ2b, I_γ2a-C_γ2a, I_ε-C_ε and I_α-C_α transcripts, and post-recombination I_μ-C_γ3, I_μ-C_γ1, I_μ-C_γ2b, I_μ-C_γ2a, I_μ-C_ε and I_μ-C_α transcripts, as normalized to *CD79b* expression. Values are mean Data are ± s.e. (bars) of two of triplicate samples. Data are from two of 3 pairs of *Hoxc4*$^{+/+}$ and *Hoxc4*$^{–/–}$ mice analyzed. Data from the third pair are depicted in **Fig. 5**.

a

pair 1

Hoxc4$^{+/+}$

	A	G	C	T
A	-	39	11	27
G	46	-	21	20
C	14	7	-	25
T	14	11	37	-

Total: 272

Hoxc4$^{-/-}$

	A	G	C	T
A	-	21	6	15
G	17	-	6	5
C	0	0	-	11
T	5	3	16	-

Total: 105

pair 2

Hoxc4$^{+/+}$

	A	G	C	T
A	-	37	17	26
G	35	-	14	13
C	9	2	-	10
T	10	12	23	-

Total: 208

Hoxc4$^{-/-}$

	A	G	C	T
A	-	19	6	9
G	12	-	5	3
C	2	3	-	12
T	7	0	10	-

Total: 88

pair 3

Hoxc4$^{+/+}$

	A	G	C	T
A	-	35	13	23
G	21	-	16	17
C	11	7	-	16
T	7	13	19	-

Total: 197

Hoxc4$^{-/-}$

	A	G	C	T
A	-	13	3	6
G	20	-	10	9
C	7	1	-	4
T	6	2	8	-

Total: 89

b

Hoxc4$^{+/+}$

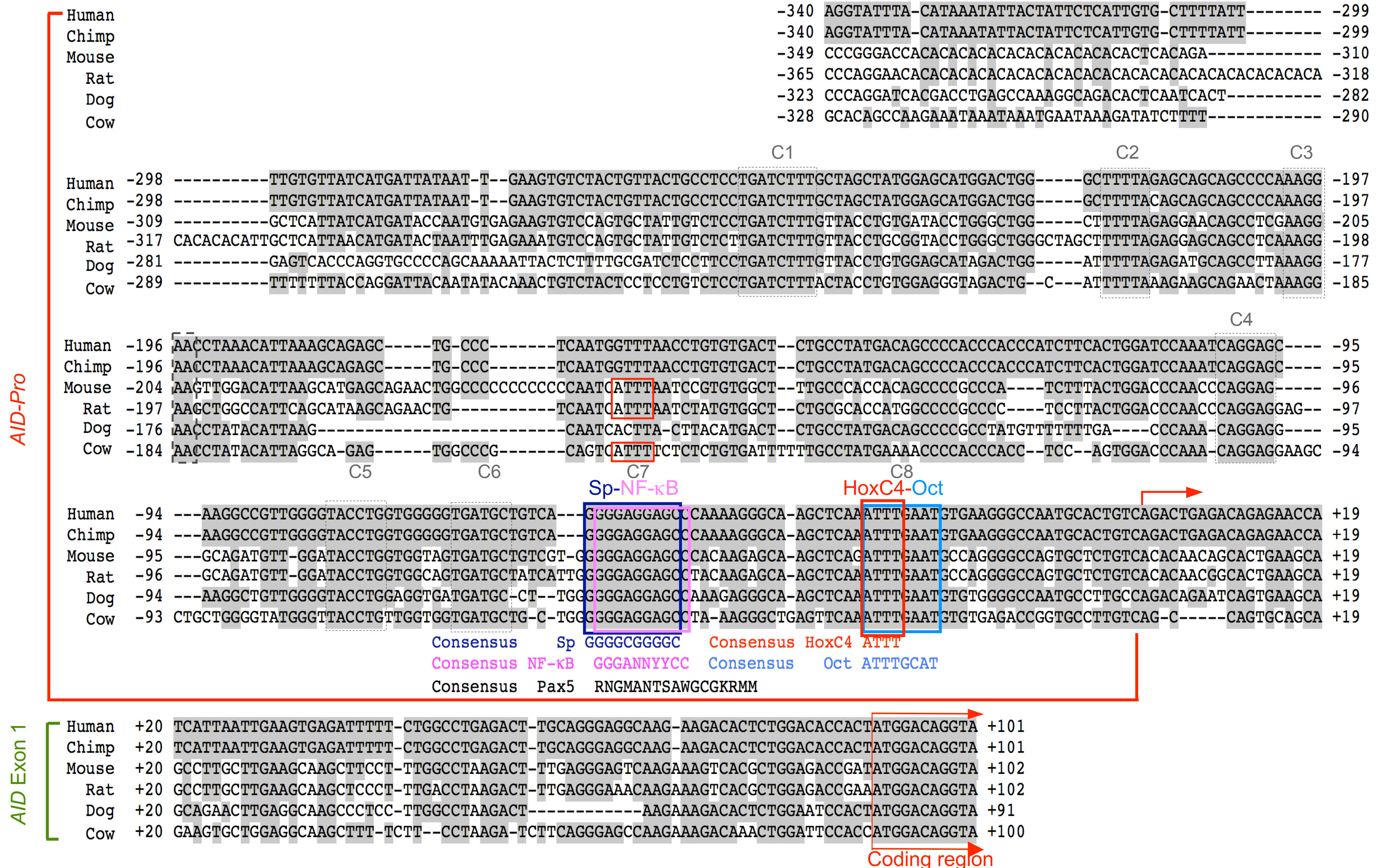
Mutations from: / to:	A	G	C	T	% total
A		16.4	6.1	11.2	33.7
G	15.3		7.6	7.5	30.4
C	5.1	2.4		7.6	15.1
T	4.6	5.4	11.8		21.8

Hoxc4$^{-/-}$

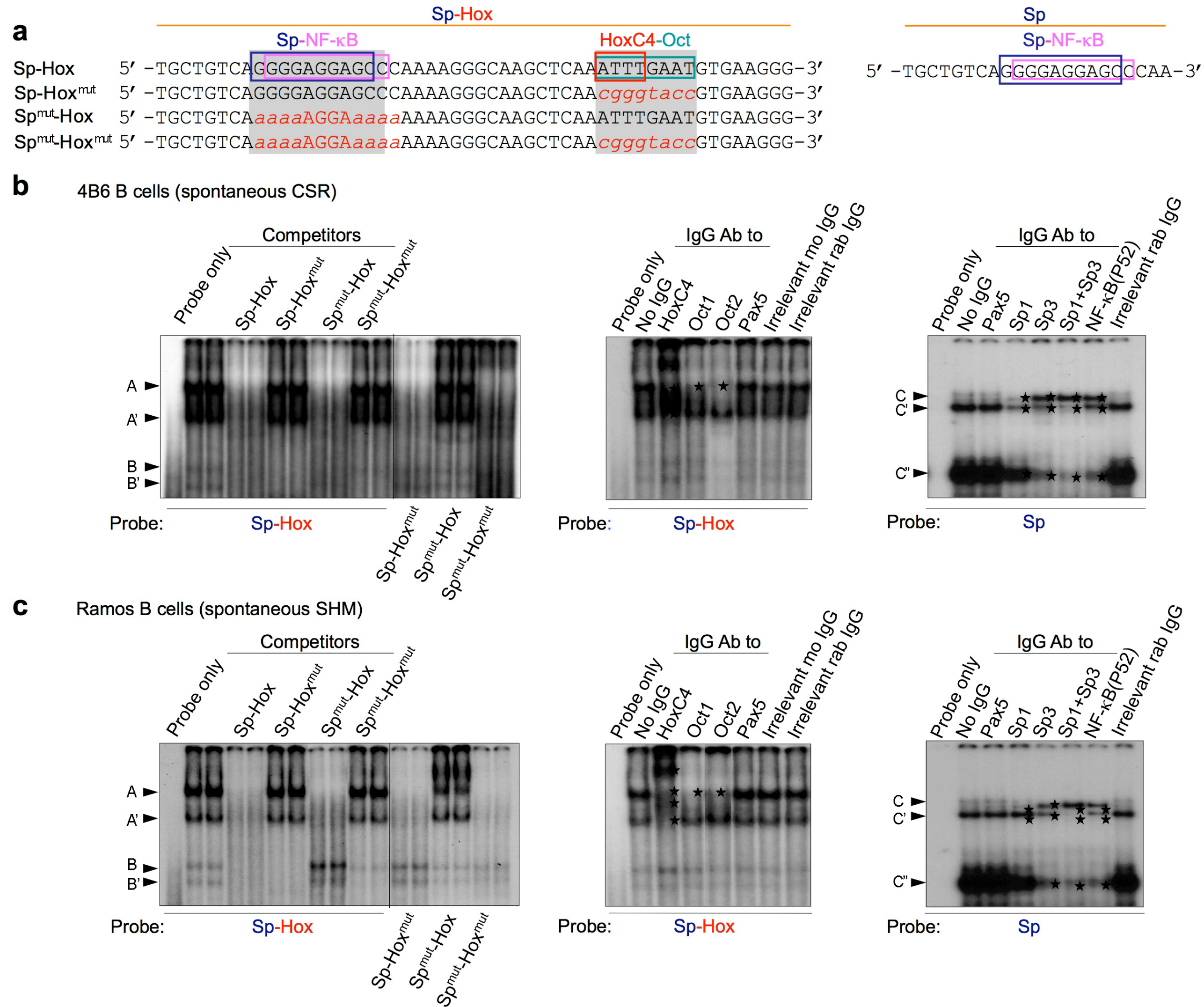
Mutations from: / to:	A	G	C	T	% total
A		18.8	5.3	10.6	34.7
G	17.4		7.4	6.0	30.9
C	3.2	1.4		9.6	14.2
T	6.4	1.8	12.1		20.2

	Hoxc4$^{+/+}$	*Hoxc4*$^{-/-}$
Mutations at dG/dC:dA/dT	44.9 : 55.1	45.0 : 55.0 (p = 0.95)
Transitions:transversions at dG/dC	50.3 : 49.7	59.8 : 40.1 (p = 0.23)
Transitions:transversions at dA/dT	50.9 : 49.1	56.1 : 43.9 (p = 0.45)
Mutation frequency	2.95×10^{-3}	1.21×10^{-3} ($p < 0.00001$)

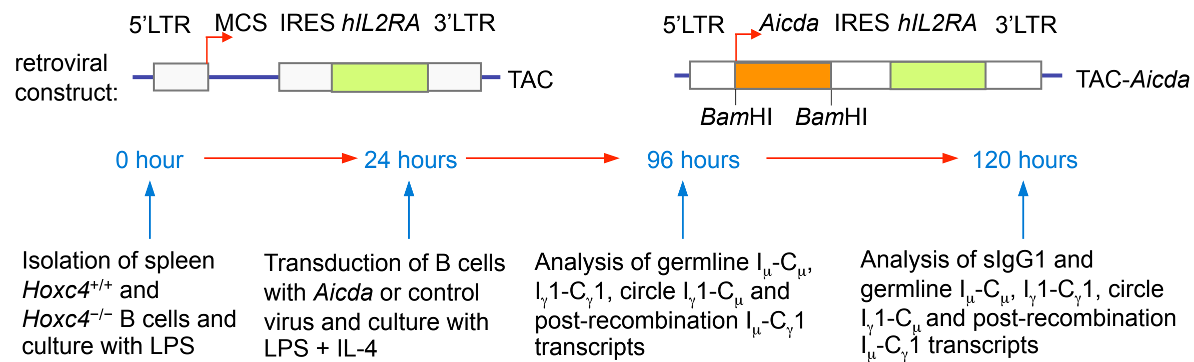
Supplementary Figure 4. Decreased somatic mutation frequency and parallel comparable reduction in mutations at dG:dC and dA:dT in the Ig H chain intronic J_H4-iEμ DNA of Peyer's patch PNA^{hi}B220$^+$ (GC) B cells from 3 12-week-old *Hoxc4*$^{-/-}$ mice as compare to their *Hoxc4*$^{+/+}$ littermates. (**a**) Numbers and nature of independent mutational events scored. (**b**) Compilations, with the numbers indicating percentages of all mutations scored in the pool of the target sequences of panels **a**. Below the compilations, the ratio of mutations at dG/dC to those at dA/dT is indicated, as is the ratio of transition:transversion substitutions at dG/dC and dA/dT.



Supplementary Figure 5. Complete sequence of the AID gene promoter (*AID-Pro*) region in the human, chimp, mouse, rat, dog and cow. The conserved regions C1-C6 did not fulfill the minimal criteria for known transcription factor-binding sites by weight matrix search using Match™ (BIOBASE Corp., Beverly, MA) (score threshold of 0.75). The conserved HoxC4-Oct-binding 5'-ATTTGAAT-3' site is boxed in red (HoxC4)-light blue (Oct); the conserved Sp-NF-κB-binding site is boxed in dark blue (Sp)-pink (NF-κB). The consensus sequences for HoxC4-, Oct-, Sp-, NF-κB- and Pax5-binding sites are depicted in alignement with the respective conserved sites (N = A, G, C or T, Y = C or T, R = A or G, M = A or C, S = G or C, W = A or T and K = G or T). Red arrow marks the putative transcription initiation site and the beginning of the coding region. Grey marks DNA sequences conserved among the six species.



Supplementary Figure 6. HoxC4, Oct1 and Oct2 bind to the conserved 5'-ATTTGAAT-3' site, and Sp1, Sp3 and NF-κB bind to the conserved 5'-GGGGAGGAGCCA-3' site in the *AICDA* promoter sequence. (**a**) Oligonucleotides containing both the HoxC4-Oct- and the Sp-NF-κB-binding sites (Sp-Hox), or the Sp-NF-κB-binding site alone (Sp) were used as probes in EMSA, and so were the mutated oligonucleotides, in which the HoxC4-Oct- and/or Sp-NF-κB-binding sites were non-functional (Sp-Hox^{mut}, Sp^{mut}-Hox and Sp^{mut}-Hox^{mut}). Nuclear proteins from spontaneously switching 4B6 B cells (**b**), and spontaneously hypermutating Ramos B cells (**c**) specifically bound to oligonucleotide probes containing the *AICDA* promoter HoxC4-Oct- and/or Sp-NF-κB-binding sites. Efficient competition was achieved by 100-fold molar excess of the unlabeled (cold) WT Sp-Hox but not cold Sp-Hox^{mut}, Sp^{mut}-Hox or Sp^{mut}-Hox^{mut} oligonucleotide. A and A' denote protein-DNA complexes involving HoxC4-Oct; B, B' and C, C', C" denote protein-DNA complexes involving Sp-NF-κB. The formation of the DNA-binding complexes was shifted or inhibited by a specific mAb to HoxC4, or Abs to Oct1, Oct2, Sp1, Sp3 or NF-κB (p52 subunit), but not by Ab to Pax5. Mouse or rabbit IgG with irrelevant binding activity served as a negative control.



Supplementary Figure 7. Top: Schematics of the control TAC and AID-expression TAC-*Aicda* retroviral constructs. The *IL2RA* gene encoding human CD25 and the *Aicda* gene encoding mouse AID are indicated. LTR, long terminal repeat; MCS, multiple cloning site. **Bottom**: protocol used to transduce *Hoxc4*$^{+/+}$ and *Hoxc4*$^{-/-}$ mouse spleen B cells.

Supplementary Table 1. Primers for real-time qRT-PCR, semiquantitative RT–PCR, amplification of intronic J_H4-iEμ DNA, *HoxC4* genotyping and ChIP assays.

	Forward primer	Reverse primer
Real-time qRT-PCR primers		
HoxC4	5′-CTACCTGACCCGAAGGAGAA-3′	5′-TGACCTCACTTTGGTGTTGG-3′
Aicda	5′-TGCTACGTGGTGAAGAGGAG-3′	5′-TCCCAGTCTGAGATGTAGCG-3′
Gapdh	5′-TTCACCACCATGGAGAAGGC-3′	5′-GGCATGGACTGTGGTCATGA-3′
CD79b	5′-CCACACTGGTGCTGTCTTCC-3′	5′-GGGCTTCCTTGGAAATTCAG-3′
Germline transcripts		
Iμ-Cμ	5’-ACCTGGGAATGTATGGTTGTGGCTT-3’	5’-TCTGAACCTTCAAGGATGCTCTTG-3’
Iγ3-Cγ3	5’-AACTACTGCTACCACCACCACCAG-3’	5’-ACCAAGGGATAGACAGATGGGG-3’
Iγ1-Cγ1	5’-TCGAGAGCCTGAGGAATGTG-3’	5’-ATGGAGTTAGTTTGGGCAGCA-3’
Iγ2b-Cγ2b	5’- GATGGGGAGGAGTTGGCAGAT-3’	5’- CGGAGGAACCAGTTGTATC-3’
Iγ2a-Cγ2a	5’-GCTGATGTACCTACCGAGAGA-3’	5’-GCTGGGCCAGGTGCTCGAGGTT-3’
Iε-Cε	5’-ACTAGAGATTCACAACG-3’	5’-AGCGATGAATGGAGTAGC-3’
Iα-Cα	5’-CAAGAAGGAGAAGGTGATTCAG-3’	5’-GAGCTGGTGGGAGTGTCAGTG-3’
Post-recombination transcripts		
Iμ-Cγ3	5’-CTCGGTGGCTTTGAAGGAAC-3’	5’-ACCAAGGGATAGACAGATGGGG-3’
Iμ-Cγ1	5’-ACCTGGGAATGTATGGTTGTGGCTT-3’	5’-ATGGAGTTAGTTTGGGCAGCA-3’
Iμ-Cγ2b	5’-ACCTGGGAATGTATGGTTGTGGCTT-3’	5’-CGGAGGAACCAGTTGTATC-3’
Iμ-Cγ2a	5’-ACCTGGGAATGTATGGTTGTGGCTT-3’	5’-GCTGGGCCAGGTGCTCGAGGTT-3’
Iμ-Cε	5’-CTCGGTGGCTTTGAAGGAAC-3’	5’-AGCGATGAATGGAGTAGC-3’
Iμ-Cα	5’-ACCTGGGAATGTATGGTTGTGGCTT-3’	5’-TAATCGTGAATCAGGCAG-3’
Mature $V_{J558}DJ_H$-Cμ transcripts	5’-AGCCTGACATCTGAGGAC-3′	5’-TGGTGCTGGGCAGGAAGT-3’
Semiquantitative RT–PCR		
HOXC4	5’-ATTCCAGCATCACCACCAGGAG-3′	5’-GGGTCAGGTAGCGGTTGTAATG-3′
AICDA	5’-TGCTCTTCCTCCGCTACATCTC-3’	5’-AACCTCARACAGGGGCAAAAGG-3’
GAPDH	5’-ACCAACTGCTTAGCACCCCT-3’	5’-CACAGTCTTCTGGGTGGCAG-3’
HoxC4	5’-TTCACGTTAGCACGGTGAAC-3’	5’-TCACTTTGGTGTTGGGGAGT-3’
Aicda	5’-GAGGGAGTCAAGAAAGTCACGCTGGA-3’	5’-GGCTGAGGTTAGGGTTCCATCTCAG-3’
Circle Iγ1-Cμ transcripts	5’-GGCCCTTCCAGATCTTTGAG-3’	5’-GAAGACATTTGGGAAGGACTGAC-3’
Gapdh	5'-ATCACTGCCACCCAGAAGACTG-3'	5'-CCCTGTTGCTGTAGCCGTATTC-3'
Intronic J_H4-iEμ DNA		
First round	5’-AGCCTGACATCTGAGGAC-3′	5’-TCTGATCGGCCATCTTGACTC-3′
Second round	5’-CATCTGAGGACTCTGCNGTCTAT-3′	5’-CCTCACTCCCATTCCTCGGTTAAA-3′
Genotyping		
HoxC4	5’-AGCAAGCAACCCATAGTCTACCC-3’	5’-ATAACCTGGTGATGTCCTCTGCCC-3’
ChIP Assay		
Human *AICDA* promoter	5’-ACTGGGCTTTTAGAGCAGCA-3’	5’-ACAGTGCATTGGCCCTTC-3’
Mouse *Aicda* promoter	5’-GGAGGCAGATGTTGGATACC-3′	5’-ATATCGGTCTCCAGCGTGAC-3′

Supplementary METHODS

Immunoblotting. B cell extracts (25 μg) were fractionated through 10% SDS-PAGE. Proteins were blotted onto polyvinylidene difluoride membranes (Bio-Rad Laboratories, Inc.) overnight (30 V) at 4 °C and then detected using primary (1:250 to 1:1000) and secondary (1:2500) Abs. After washing with PBS-Tween 20 (0.05%), bound HRP-conjugated Abs or mAbs were detected using Western Lightning® Plus–Enhanced Chemiluminescence reagents (PerkinElmer Life and Analytical Sciences, Inc.).

NP_{16}-CGG immunization and titration of NP-binding IgM and IgG1. *Hoxc4*$^{-/-}$ and *Hoxc4*$^{+/+}$ C57BL/6 mice (8- to 10-week-old) that were free of obvious disease were given a first intraperitoneal injection of 100 μg of NP-CGG at a ratio of 16:1 (NP_{16}-CGG) (Biosearch Technologies, Inc.) in Imject® alum (Pierce). After 21 d, they were boosted intraperitoneally with another 100 μg NP_{16}-CGG in alum. Blood was collected 28 d after the primary immunization for titration of total and NP-binding IgM and IgG1 using specific ELISAs. To measure total IgM and IgG1, mouse sera were serially twofold diluted starting from 1:10,000 (for IgM) or 1:40,000 (for IgG1), and 100 μl of these were added to each of the wells and incubated at 37 °C for 1 h. After washing the plates, biotin-labeled anti-IgM or anti-IgG1 mAb was added to the wells for 1 h, and then detected using horseradish peroxidase-streptavidin, followed by OPD. The reaction was stopped with sulphuric acid after 30 min., before measuring O.D. at 492 nm. For NP_{30}- or NP_3-binding IgM and IgG1 titers, 96 well plates were coated with NP_{30}- or NP_3-BSA. Mouse sera were 1:100 then serially twofold diluted before being applied to 96-well plates. Titers were expressed in the 50% effective concentration (EC_{50}) units. An EC_{50} was defined as the number of dilutions needed to reach 50% of the saturation binding, as calculated using GraphPad Prism® software (GraphPad Software, Inc.). All assays were performed in triplicates.

Histology. Spleens from NP_{16}-CGG immunized mice were embedded in OCT compound, snap-frozen and stored at –80 °C. Cryostat sections (7 μm) were fixed using cold acetone and stored at –80 °C for 25 min and then air-dried for 30 min at 25 °C. Phycoerythrin (PE)-labeled anti-mouse B220 mAb (clone RA3-6B2) eBioscience Corp.) (1:200 dilution) and fluorescein isothiocyanate (FITC)-labeled peanut agglutinin (PNA) (E-Y Laboratories, Inc.) (1:100 dilution) were applied to the sections, which were kept in a dark box at 25 °C for 1 h. After washing with PBS, the sections were mounted using anti-fade reagent (Invitrogen Corp.) for examination.

EMSA and EMSA shift assays. Nuclear extracts from B cells were prepared using a microprocedure involving hypotonic lysis followed by high salt nuclei extraction[1-3]. Oligonucleotides encompassing the HoxC4-Oct and/or Sp-NF-κB *cis*-elements in *AICDA* promoter were as follows: HoxC4-Oct-Sp-NF-κB, 5′-TGCTGTCAGGGGAGGAGCCCAAAAGGGCAAGCTCAAATTTGAATGTGAAGGG-3′ (Sp-Hox), 5′-TGCTGTCAGGGGAGGAGCCCAAAAGGGCAAGCTCAA*cgggtacc*GTGAAGGG-3′ (Sp-Hox^{mut}), 5′-TGCTGTCA*aaaa*AGGA*aaaa*AAAAGGGCAAGCTCAAATTTGAATGTGAAGGG-3′ (Sp^{mut}-Hox), 5′-TGCTGTCA*aaaa*AGGA*aaaa*AAAAGGGCAAGCTCAA*cgggtacc*GTGAAGGG-3′ (Sp^{mut}-Hox^{mut}) and Sp-

NF-κB, 5′-TGCTGTCAGGGGAGGAGCCCAA-3' (Sp). All reactions were performed as reported[1-3]. For supershift-inhibition EMSA reactions, 5 μg of anti-HoxC4 mouse mAb (clone 1E9) (Novus Biologicals, Inc.), anti-Oct1, anti-Oct2, anti-OcaB, anti-Pax5, anti-Sp1, anti-Sp3 or anti-NF-κB (p52 subunit) rabbit polyclonal Abs (Santa Cruz Biotechnology, Inc.) were pre-incubated with nuclear extracts for 30 min prior to addition of probe. All EMSA gels were 7% polyacrylamide and subjected to electrophoresis in 0.25x TBE buffer, pH 7.5. Gels were dried and exposed for autoradiography.

REFERENCES

1. Schaffer, A., Cerutti, A., Shah, S., Zan, H. & Casali, P. The evolutionary conserved sequence upstream of the human Ig Sγ3 region is an inducible promoter: Synergistic activation by CD40 ligand and IL-4 via cooperative NF-κB and STAT-6 binding sites. *J. Immunol.* **162**, 5327-5336 (1999).
2. Schaffer, A. *et al.* Selective inhibition of class switching to IgG and IgE by recruitment of the HoxC4 and Oct-1 homeodomain proteins and Ku70/Ku86 to newly identified ATTT cis-elements. *J. Biol. Chem.* **278**, 23141-23150 (2003).
3. Kim, E.C., Edmonston, C.R., Wu, X., Schaffer, A. & Casali, P. The HoxC4 homeodomain protein mediates activation of the immunoglobulin heavy chain 3' hs1,2 enhancer in human B cells. Relevance to class switch DNA recombination. *J. Biol. Chem.* **279**, 42258-42269 (2004).