ONLINE METHODS

ChIP-seq. Resting splenic B cells were isolated from 6- to 8-week-old wild-type C57BL6/J or *Aicda*$^{-/-}$ mice (backcrossed for 12 generations to C57BL6/J mice) with anti-CD43 Microbeads (anti-Ly48; Miltenyi Biotech) and were cultured for 60 h with LPS (50 μg/ml; Sigma) and IL-4 (5 ng/ml; Sigma). In total, 57 mice were used for ChIP-seq studies. For quality control, switch recombination to IgG1 was checked before deep-sequencing analysis (**Supplementary Fig. 9**). Chromatin was digested by incubation of collected cells with micrococcal nuclease (1 unit/ml; Sigma) for analysis of most histone modifications except H3K79me1, H3K79me2 and H3K79me3. For those last three modifications, as well as PolII, p300, CTCF and AID, cells were crosslinked for 10 min at 37 °C with 1% (vol/vol) formaldehyde, followed by quenching with 0.125 M glycine (final concentration). Crosslinked cell samples were then sonicated to obtain DNA fragments 200–500 bp in length. For each ChIP sample, sonicated (or nuclease-treated) chromatin (from 1×10^7 cells) was precleared with 40 μl protein A or protein G magnetic beads, followed by incubation overnight at 4 °C in radioimmunoprecipitation assay (RIPA) buffer (10 mM Tris, pH 7.6, 1 mM EDTA, 0.1% (wt/vol) SDS, 0.1% (wt/vol) sodium deoxycholate and 1% (vol/vol) Triton X-100) with 40 μl protein A or G magnetic beads (Invitrogen) and 5–10 μg antibody (**Supplementary Table 7**). Beads were washed twice with RIPA buffer alone, then twice with RIPA buffer plus 0.3 M NaCl, twice with LiCl buffer (0.25 M LiCl, 0.5% (vol/vol) Igepal CA630 and 0.5% (wt/vol) sodium deoxycholate), once with Tris-EDTA buffer (Tris 10 mM, pH 8.0, and 1 mM EDTA) plus 0.2% (vol/vol) Triton X-100, and once with Tris-EDTA buffer alone. ChIP DNA was extracted for 4 h at 65 °C in Tris-EDTA buffer with 0.3% (wt/vol) SDS and proteinase K (1 mg/ml). Samples were sequenced on a Genome Analyzer IIx after being prepared for microsequencing according to the manufacturer's protocol (Illumina). Each genome-wide analysis represents a single immunoprecipitation analyzed in one lane of the Ilumina slide. The number of immunoprecipitations is as follows: AID, four (including an *Aicda*$^{-/-}$ control); RPA, ten (details, **Supplementary Table 5**); PolII, four; H3K4me3, H3K4me27, H2BK5Ac, H3K27Ac, H3K9Ac, H2BK120Ac and CTCF, three; other samples, one. Although the Ab1 antibody to AID has been used for ChIP analysis before[22], its immunoprecipitation specificity was further tested by immunoblot analysis (**Supplementary Fig. 10**); all experiments used Ab1 except those in **Supplementary Figure 2**.

Hypermutation analysis. CD43$^-$ splenocytes from *Igk*-AID *Ung*$^{-/-}$ or *Aicda*$^{-/-}$ mice (at a density of 0.1×10^6 cells per ml) were cultured for 120 h with LPS and IL-4, plus anti-CD180 (anti-RP105; 0.5 μg/ml; RP/14; BD Pharmingen). Cells were passaged at 72 h at a dilution of 1–4. Genomic DNA (50 ng) was amplified for 30 cycles with Phusion DNA polymerase (New England Biolabs) and specific primers (**Supplementary Table 8**). For nested PCR, two 20-cycle amplifications were done. The amplicon was cloned with a Zero Blunt PCR kit (Invitrogen) and sequenced.

Mice. All animal experiments were done according to the National Institutes of Health guidelines for laboratory animals and were approved by the Scientific Committee of the National Institute of Arthritis, Musculoskeletal and Skin Diseases and National Cancer Institute Animal Facilities.

Sequencing-based mRNA analysis. The mRNA from activated B cells was sequenced according to the manufacturer's protocol for deep-sequencing analysis of mRNA (Illumina). A detailed version of this protocol has been published[34].

Image processing and alignment. Standard Illumina image-analysis and base-calling software was used for determination of short-read sequences and corresponding qualities as well as application of the default quality filters. Pipeline output was transformed to the fastq format and reads were aligned to the Build 37 assembly of the National Center for Biotechnology Information mouse genome data (July 2007; NCBI37/mm9) with Bowtie 0.9.9.3 software, including the following command line options: –best–all -m1 -n2 (reads that align uniquely in the best alignment stratum, allowing up to two mismatches).

ChIP-seq analysis. Peaks of enrichment, assessed relative to those of a random background model, were identified with SICER 1.03 software (spatial clustering approach for the identification of ChIP-enriched regions). For AID, the parameters used were as follows: window size, 100; gap size, 100; e, 0.000001; redundancy, 1; fragment size, 127 (*Aicda*$^{+/+}$) or 115 (*Aicda*$^{-/-}$). Estimates of fragment size were calculated from the data on the basis of a shifting algorithm that attempts to minimize a Shannon entropy measure. For other modifications, a more lenient 'e' value of 0.05 was used. Read counts were adjusted for library size and are presented as reads per million nonredundant aligned reads. Density plots presented here include only windows that were on enriched peak regions. A combination of custom Python and R software was used for downstream analysis of SICER output (for example, the assignment of individual reads to gene regions and calculation of cumulative summary plots at TSSs).

AID ChIP-seq saturation analysis. Multiple lanes in multiple flow cells were combined for *Aicda*$^{+/+}$ samples (eight lanes; 143,995,816 reads) and *Aicda*$^{-/-}$ samples (six lanes; 134,418,212 reads) for ChIP-seq with anti-AID. For saturation analysis, each data set was split into randomly generated subsets of approximately 6×10^6 reads, then increasing numbers of subsets were merged, and each set generated in this way was analyzed separately as described above.

Enrichment of ChIP-seq tags on AID islands. The total number of aligned reads of a ChIP-seq experiment was used to estimate the expected number of such reads that would be on AID islands by chance and was compared with the actual number of reads on the islands. The significance of the enrichment as well as the 99% confidence interval was determined with a two-tailed binomial test.

Hypermutation *q* values. The following procedure was used for the adjustment of *P* values to control for false-discovery rate. Fisher's exact test was used to test the significance of the difference between *Aicda*$^{+/+}$ and *Aicda*$^{-/-}$ samples in their mutation rate. To control for false discovery rate, the *P* values were further adjusted with the p.adjust function in R software.

ChIP-seq of small RNA (divergent transcription analysis). On the basis of previous ChIP-seq analysis of small RNA, tags that aligned uniquely with the genome and either mapped to protein-coding domains or could not be mapped to any annotated regions were retrieved from the read database for activated B cells (three experiments) and germinal center cells (one experiment). Up to four identical tags were retained and mapped to the RefSeq annotation. Approximately 1 million tags could be assigned to a known transcript, and 0.1 million could be assigned to a putative divergent transcript on the opposite strand of a known transcript up to –2 kb upstream of the TSS.

Hierarchical clustering. The Partek Genomics Suite version 6.5 was used for hierarchical clustering, with $\log_2$-transformed data and normalized Euclidean distance (average Euclidean distance as implemented in Partek). A distance matrix calculated from the densities of each modification and AID on the regions with significant AID tag densities was used for hierarchical clustering.