Science Immunology
AAAS

immunology.sciencemag.org/cgi/content/full/4/34/eaau7523/DC1

Supplementary Materials for

TET enzymes augment activation-induced deaminase (AID) expression via 5-hydroxymethylcytosine modifications at the *Aicda* superenhancer

Chan-Wang J. Lio*, Vipul Shukla, Daniela Samaniego-Castruita, Edahi González-Avalos, Abhijit Chakraborty, Xiaojing Yue, David G. Schatz, Ferhat Ay, Anjana Rao*

*Corresponding author. Email: arao@lji.org (A.R.); lio@lji.org (C.-W.J.L.)

Published 26 April 2019, *Sci. Immunol.* **4**, eaau7523 (2018)
DOI: 10.1126/sciimmunol.aau7523

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at immunology.sciencemag.org/cgi/content/full/4/34/eaau7523/DC1)

Materials and Methods --- Bioinformatics analyses
The reference genome used was mm10. Heatmaps and profile plots were generated using DeepTools(*59*).

CMSIP analysis
Paired-end reads (50bp) were mapped to the mouse genome mm10 GRCm38 (Dec. 2011) from UCSC, using BSMAP (V.2.74) (-v 4 -R -n 1 -w 2 -r 0 -q 20 -R -p 8)(*60*). Reads that mapped to the spike-in control (Lambda and Puro) were filtered out from the Sam file using awk. Tag directories were created with the remaining reads using makeTagDirectory from HOMER(*61*) (-genome mm10 -tbp 1 –checkGC). Peaks were called with findPeaks from HOMER (-style histone -o auto -i). Peaks from all samples were merged with mergePeaks from HOMER into a master table. Quantile normalization was applied to all raw counts files and differentially enriched 5hmC regions were identified with edgeR(*62*); a *p* adjusted value of ≤ 0.05 was used as a cutoff.

H3K27Ac ChIP analysis
Single end reads (50bp) were mapped to the mouse genome mm10 GRCm38 (Dec. 2011) from UCSC with Bowtie (V.1.1.2). Reads were sorted and PCR duplicates were removed using SortSam and MarkDuplicates, respectively from Picard Tools (V.2.7.1). Tag directories were created with makeTagDirectory (-genome mm10 –checkGC) from HOMER, and peaks were called with findPeaks (-region). Peaks from all samples were merged with mergePeaks from HOMER into a master table. Quantile normalization was applied to all raw counts files and differentially enriched 5hmC regions were identified with edgeR(*62*); a *p* adjusted value of ≤ 0.05 was used as a cutoff.

BATF ChIP analysis
Single end reads (50bp) were mapped to the mouse genome mm10 GRCm38 (Dec. 2011) from UCSC with Bowtie (v1.1.2). Reads were sorted and PCR duplicates were removed using SortSam and MarkDuplicates, respectively from Picard Tools (V.2.7.1). Tag directories were created with makeTagDirectory (-genome mm10 –checkGC) from HOMER(*61*), and peaks were called with findPeaks (-style factor -o auto).

Definition of preferentially active enhancers
Preferentially active enhancers (**Fig. S1H**) were defined as distal H3K27Ac regions (> 1kb from TSS) that had ATAC-seq and H3K27Ac peaks overlapping in at least 50% of either peak/region; overlapping was calculated with intersectBed -f 0.5 -f 0.5 –e (Bedtools v2.26.0). The differential enrichment in an enhancer was called if it contains a differentially enriched H3K27Ac region as well as at least one differentially access region.

ATAC-seq analysis
Paired-end reads (100 bp) were mapped to the mouse genome mm10 GRCm38 (Dec. 2011) from UCSC using Bowtie 1.0.0 ("-p 8 -m 1 --best --strata -X 2000 -S --fr --chunkmbs 1024."). Reads that failed this alignment step were filtered for Illumina adapters and low quality using “Trim Galore!” ("--paired --nextera --length 37 --stringency 3 --three_prime_clip_R1 1 --three_prime_clip_R2 1") and re-mapped using the same parameters. Both mapping results were merged and processed together to remove duplicates using picard-tools-1.94 MarkDuplicates. Mitochondrial and Chromosome Y reads were excluded.
Subnucleosomal fragments were obtained with SAMtools and awk to identify DNA fragments that were less than or equal to 100 nt in length. These fragments were used to call peaks using HOMER (v4.9.1) findPeaks function for each replicate ("-size 500 -region -center -P .1 -LP .1 -poisson .1 -style dnase") and all the peak sets were merged to generate a global set. Peaks overlapping with ENCODE blacklisted regions(*63*) were removed. From each sample, Tn5 insertion sites were obtained by isolation of the initial 9bp of mapped reads(*64*) which were used to compute the number of transposase insertions per peak using MEDIPS(*65*). Raw reads from all samples were quantile-normalized prior to differential coverage analysis using edgeR without TMM (Trimmed mean of M-values) normalization. Only regions with more than 32 normalized reads across the samples per comparison. Differentially accessible regions were defined by an adjusted *p* value (FDR) lower than 0.05 and a log 2 fold enrichment higher equal than 1.

OxBS analysis
OxBS-seq reads were mapped to both the mouse genome mm10 GRCm38 (Dec. 2011) from UCSC and the phage Lambda genome (GenBank: J02459.1) using bsmap-2.90 (" -v 15 -w 3 -p 4 -S 1921 -q 20 -A AGATCGGAAGAGC -r 0 -R -V 2 "). The mapping results were separated into reads belonging to the mm10

genome and each of the three loci from lambda used for oxidation and conversion efficiency calculation. Methylation calls from lambda- and mm10-derived reads were obtained using bsmap-2.90 function methratio.py (" -u -p -g -i "correct" -x CG,CHG,CHH "). Conversion efficiencies as well as posterior probabilities of methylation, hydroxymethylation and unmodified cytosine were calculated by luxGLM v.0.666 (prior probabilities used for for C, hmC and mC "998,1,1", "6,2,72" and "1,998,1" respectively)(*66*). Following genomic positions from lambda used for oxidation and BS treatment efficiencies: chrLambda:22893-23053 C; chrLambda:23765-23925 hmC; chrLambda:47335-47495 mC.

WGBS analysis

WGBS reads were mapped to both the mouse genome mm10 GRCm38 (Dec. 2011) from UCSC. Bisulfite conversion efficiency was estimated based on cytosine methylation in non-CpG context. For all the samples the bisulfite conversion efficiency was higher than 0.9996. Duplicated reads caused by PCR amplification were removed by BSeQC (v1.2.0) applying a *p* value cutoff Poisson distribution test in removing duplicate reads (1e-5)(*67*). Consequently, a maximum of three stacked reads at the same genomic location were allowed and kept for further analysis. In addition, BSeQC was employed for removing DNA methylation artifacts introduced by end repair during adaptor ligation. Overlapping segments of two mates of a pair were reduced to only one copy to avoid considering the same region twice during the subsequent DNA methylation quantification. To estimate CpG DNA methylation at both DNA strands, methratio.py script was executed from BSMAP (v2.90) (-u -r -z -g -i "correct" -x CG). To identify differentially methylated cytosines and regions (DMCs and DMRs), a naïve B cells dataset and was compared to a activated B cell replicate using RADmeth methpipe-3.4.2 (adjust -bins 1:100:1 ; merge -p 0.05)(*68*).

RNA-seq analysis

RNA-seq samples at four different time-points collected from WT and DKO conditions were first mapped to the mouse genome mm10/GRCm38 using both Hisat2(*69*) (“--no-mixed --no-discordant --add-chrname –dta”) and Tophat2(*70*) (“--no-novel-juncs”) alignment programs separately. Aligned bam files obtained from both the programs were further used to generate the Hisat2- and Tophat2-specific counts using HTseq-count program(*71*) (default parameters). Hisat2- and Tophat2-specific count files at each time point for WT and DKO conditions were then used to identify the differentially expressed genes (FDR < 0.05) between matching time points using edgeR program(*62*). Potential batch effects were removed using svaseq program(*72*). Finally, the common differentially expressed genes obtained from both Hisat2- and Tophat2-specific list were used to perform the downstream analysis.

Genome-browser track generation for ChIP-seq

ChIP-seq results from TET2, Ig control, E2A, PU.1, p300, and GCN5 were processed as follow to generate the genome browser tracks. Fastq files were mapped to mm10 reference genome with Bowtie 2 (v2.1.0) with “-very-sensitive”. The mapped SAM files were converted to BAM using Samtools (v1.7) view –h –F 4, and duplicates were removed using Picard (v2.7.1). BigWig files were made by first generating a BedGraph files from the filtered Bam files using Bedtools (v2.26.0) genomecov followed by bedGraphToBigWig (v4) with read counts normalized to 10,000,000 reads.

Miscellaneous analyses of regions

For distance between regions to the closest TSS was analyzed with HOMER software with “annotatePeaks.pl –annStats”. Overlap between regions was analyzed by “bedtools intersect” with no requirement for the degree of overlapping. The degree of significance for overlap between superenhancers and test regions was estimated by Fisher exact test using “bedtools fisher”.

Time-series analysis

For a unified analysis of the RNA-seq time-course data (0hr to 96hr) from WT samples, TC-seq package(*73*) was used on the combined RNA-seq read counts, obtained after applying Tophat2(*70*) and Hisat2(*69*) alignment programs (Described in the previous RNA-seq analysis part). TC-seq utilizes GLM method implemented in edgeR package(*62*) to detect the differential events in gene expression. Differential analysis was performed between "0hr" to the rest of the time points, and the significant differential events were extracted whenever a log2FC > 2 or < -2 and FDR < 0.05 criteria was satisfied. To detect the temporal pattern of the differential gene expressions (RPKM values), a soft clustering algorithm implemented in TCseq program

was applied ("algo = 'cm', k = 6, standardize = TRUE"). Finally, the differential genes were assigned to a cluster (C1- C6) representing a specific temporal pattern of expression, if the membership probability of the genes to a cluster is 0.5 or more.

Published datasets

Naïve H3K4me1 (0h): SRR1535686, SRR1535685. Activated H3K4me1 (48h): SRR1014530. SRR1087900. Naive WGBS (0h): SRR1003257. Activated WGBS (48h): SRR1020523. Naive PU1 (0h): SRR2976278. Activated PU1 (48h): SRR1014532. Naïve DSG control (0h): SRR3158132. Activated E2A DSG anti-CD40/IL-4 (48h): SRR3158146. Naïve Brg1 (0h): SRR3619348. Naïve Chd4 (0h): SRR3619349. Naïve Gcn5 (0h): SRR3619350. Naïve p300 (0h): SRR3619356. Activated Brg1 (24h): SRR3619334. Activated Chd4 (24h): SRR3619335. Activated Gcn5 (24h): SRR3619336. Activated p300 (24h): SRR3619342.

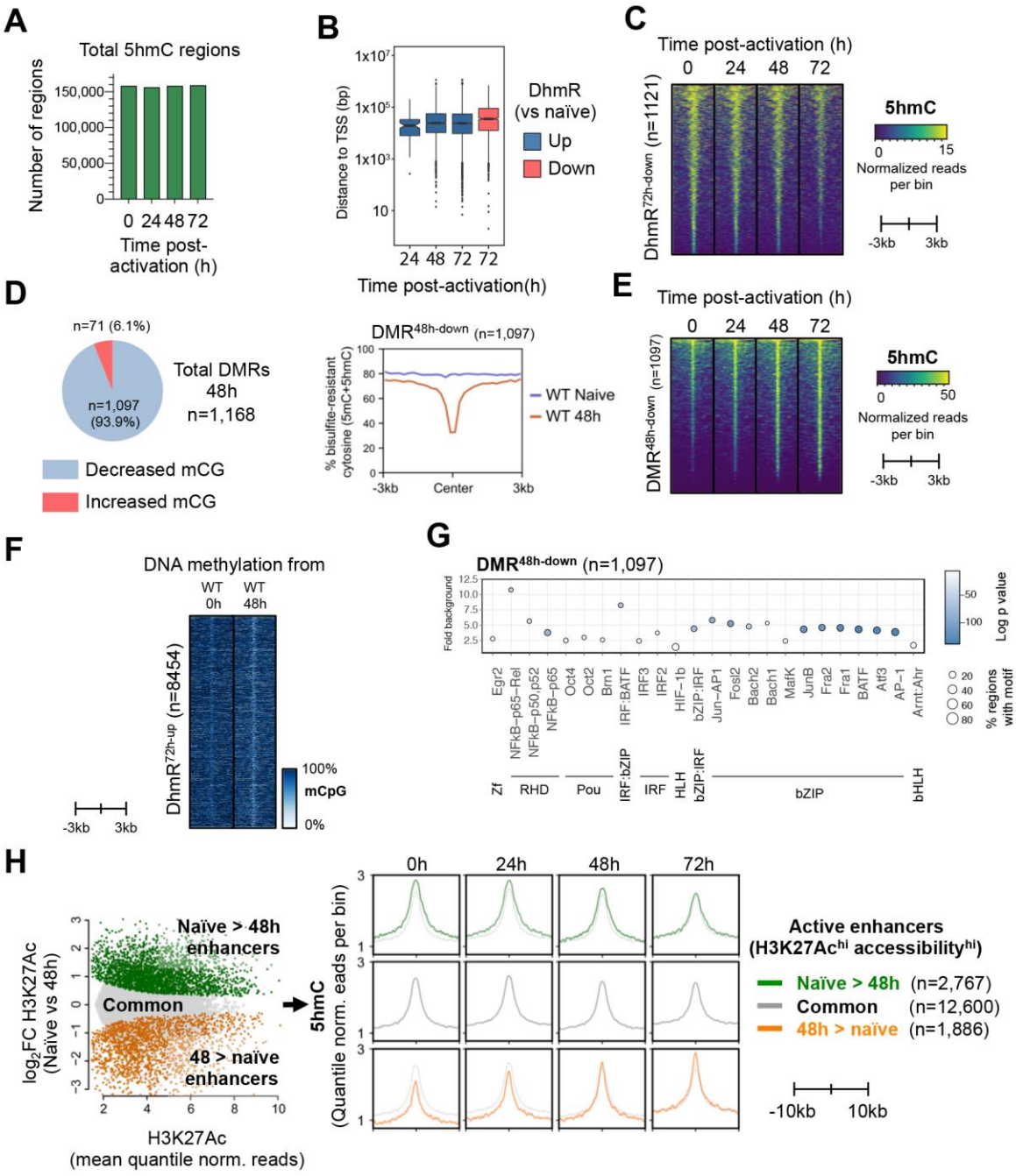

Figure S1. TET-mediated DNA hydroxymethylation correlates with demethylation and enhancer activity.
(A) Similar total numbers of 5hmC-enriched regions between naïve and activated B cells. **(B)** Box-and-whisker plot showing that differentially hydroxymethylated regions (DhmRs) in activated vs naïve B cells (see **Fig. 1C**) are located on average more than 10 kb from the closest transcription start site (TSS). **(C)** Heatmaps showing the kinetics of 5hmC modification at the 1,121 regions with decreased 5hmC in 72h-activated vs naïve B cells ($DhmR^{72h\text{-}down}$; see **Fig. 1C**). 5hmC enrichment is shown as normalized reads per 100 bp bin. **(D)** *Left*, the vast majority of differentially methylated regions (DMRs) with altered WGBS signal (5mC+5hmC) in naïve vs 48h-activated B cells show decreased DNA methylation. *Right*, plot of average DNA methylation (bisulfite-resistant cytosine 5mC+5hmC) at the $DMR^{48h\text{-}down}$ (n=1,097) in 48h-activated vs naïve B cells. Average methylation is measured per 200 bp bin. **(E)** Heatmaps showing the kinetics of change (increase) in 5hmC at the 1,097 $DMR^{48h\text{-}down}$ with decreased methylation 48h post-activation. 5hmC enrichment is shown in normalized reads per 100 bp bin. **(F)** Heatmap showing decreased DNA methylation at the 8454 $DhmR^{72h\text{-}up}$ regions. **(G)** Strong enrichment for consensus binding motifs of NFkB, IRF:bZIP, bZIP, and other transcription factors in the 1,097 regions that became demethylated after 48h of activation ($DMR^{48h\text{-}down}$). Random genomic regions were used as background for motif analysis. Y-axis indicates the fold enrichment versus background, circle size indicates the percentage of regions containing the respective motif, and the color indicates the significance (Log_{10} *p* vaule). **(H)** 5hmC levels track with enhancer activity. *Left*, MA plot showing differentially active enhancers between naïve and 48h-activated B cells were classified based on the significant difference in H3K27Ac and accessibility (ATAC-seq) into enhancers preferentially active in naïve B cells (green, "Naïve>48h") and enhancers preferentially active in 48 h-activated B cells (orange, "48>naïve"). The remaining enhancers not meeting the above criteria were classified as common active enhancers (grey, "Common"). *Right*, mean level of 5hmC per bin (50 bp) at the +/- 10kb interval to the center was plotted for each type of active enhancer. Note that the 5hmC levels from the "Common" enhancers (*middle)* are also plotted as dotted lines for naïve-B-active (*top*) and activated-B-active enhancers (*bottom*) as reference.

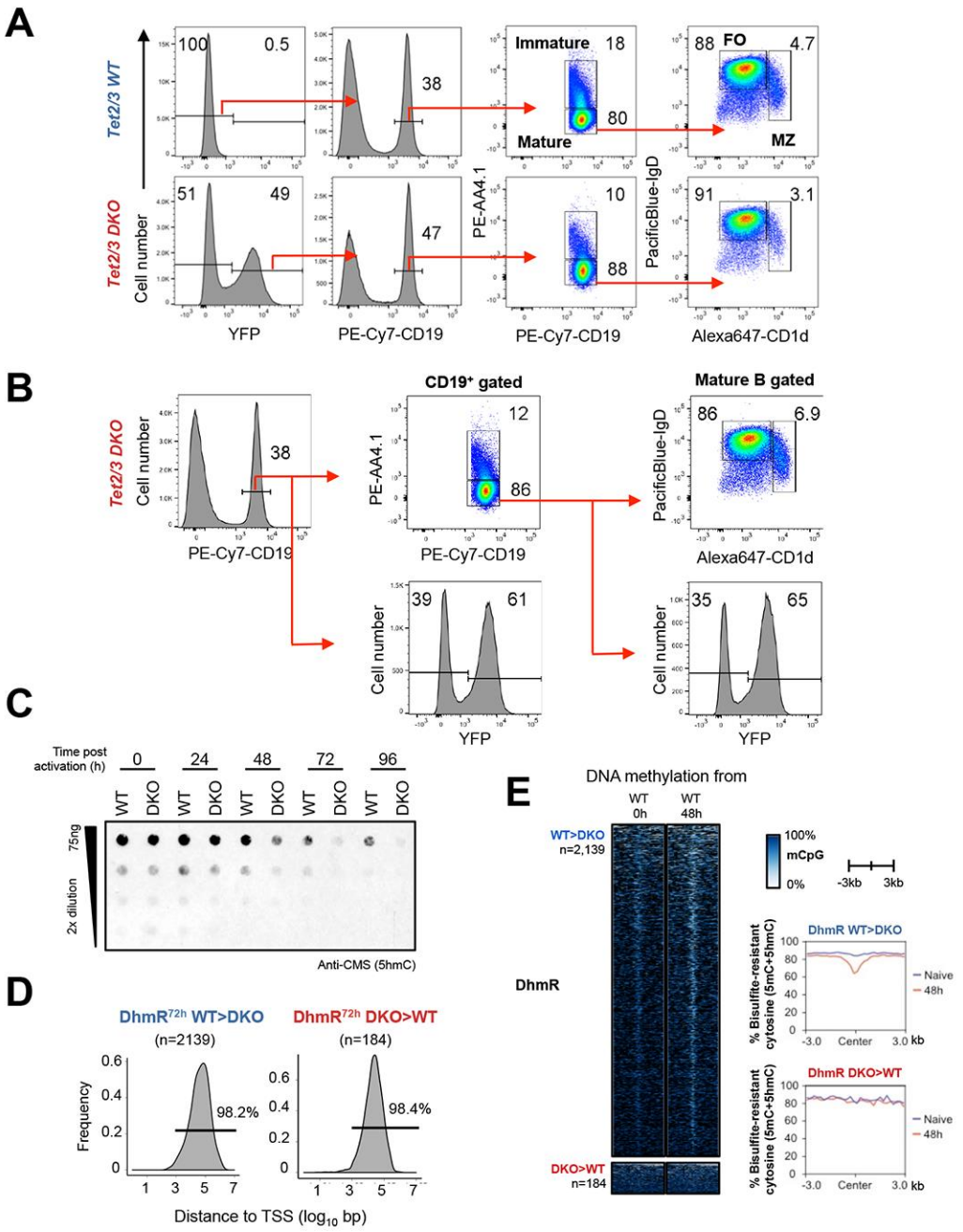

Figure S2. Phenotypic features of WT and *Tet2/3* DKO B cells.

(A) Comparable splenic mature B cell populations in *Tet2/3*-conditionally deleted mice. WT (*Tet2/3-flox Rosa26-LSL-YFP*) and DKO (*Cre*^{ERT2} *Tet2/3-flox Rosa26-LSL-YFP*) mice were treated as in **Fig. 2B** and the phenotype of splenic B cells was analyzed on day 7 after the initial tamoxifen injection. Plots were first gated on live single cells based on FSC/SSC (*first panel*) and total (WT) or YFP^{+} (DKO) CD19^{+} B cells were subsequently gated (*second panel*), followed by analysis of mature and immature B cells (*third panel*); and follicular (FO) and marginal zone (MZ) B cells (fourth panel). **(B)** Similar percentages of YFP^{+} cells in total (CD19^{+}; middle panel) and mature (CD19^{+} AA4.1^{lo}) B cells. **(C)** Total 5hmC levels in WT and *Tet2/3* DKO B cells assessed by cytosine 5-methylenesulphonate (CMS) dot blot (see *Materials and Methods*). Note that 5hmC levels decrease in *Tet2/3* DKO B cells only after several rounds of cell division (>48h). **(D)** Histograms showing the distance from the TSS to the TET-regulated DhmR regions differentially marked with 5hmC in 72h-activated *Tet2/3* DKO relative to WT B cells (see **Fig. 2D, 2E**). The 2,139 and 184 DhmRs with decreased (*left*, DhmR^{72h} WT>DKO) and increased (*right*, DhmR^{72h} DKO>WT) 5hmC after activated for 72h 72h-activated *Tet2/3* DKO relative to WT B cells are located on average more than 10 kb from the closest TSS. **(E) TET-mediated 5hmC modifications mark DNA demethylation.** *Left panels*, heatmaps show DNA methylation status in naïve and 48h-activated WT B cells (WGBS, 5mC+5hmC) at the 2,139 and 184 DhmR regions with decreased (*top*, WT>DKO) and increased (*bottom*, DKO>WT) 5hmC in 72h-activated *Tet2/3* DKO vs WT B cells. *Right panels*, plots of the average decrease in bisulfite-resistant modifications (5mC+5hmC) per bin (200 bp) at these regions. The majority of the 2,139 WT>DKO DhmRs at 72h show decreased DNA methylation in activated WT B cells (*top*); the 184 DKO>WT DhmR at 72h are fully methylated and unchanged in *Tet2/3* DKO vs WT (*bottom*).

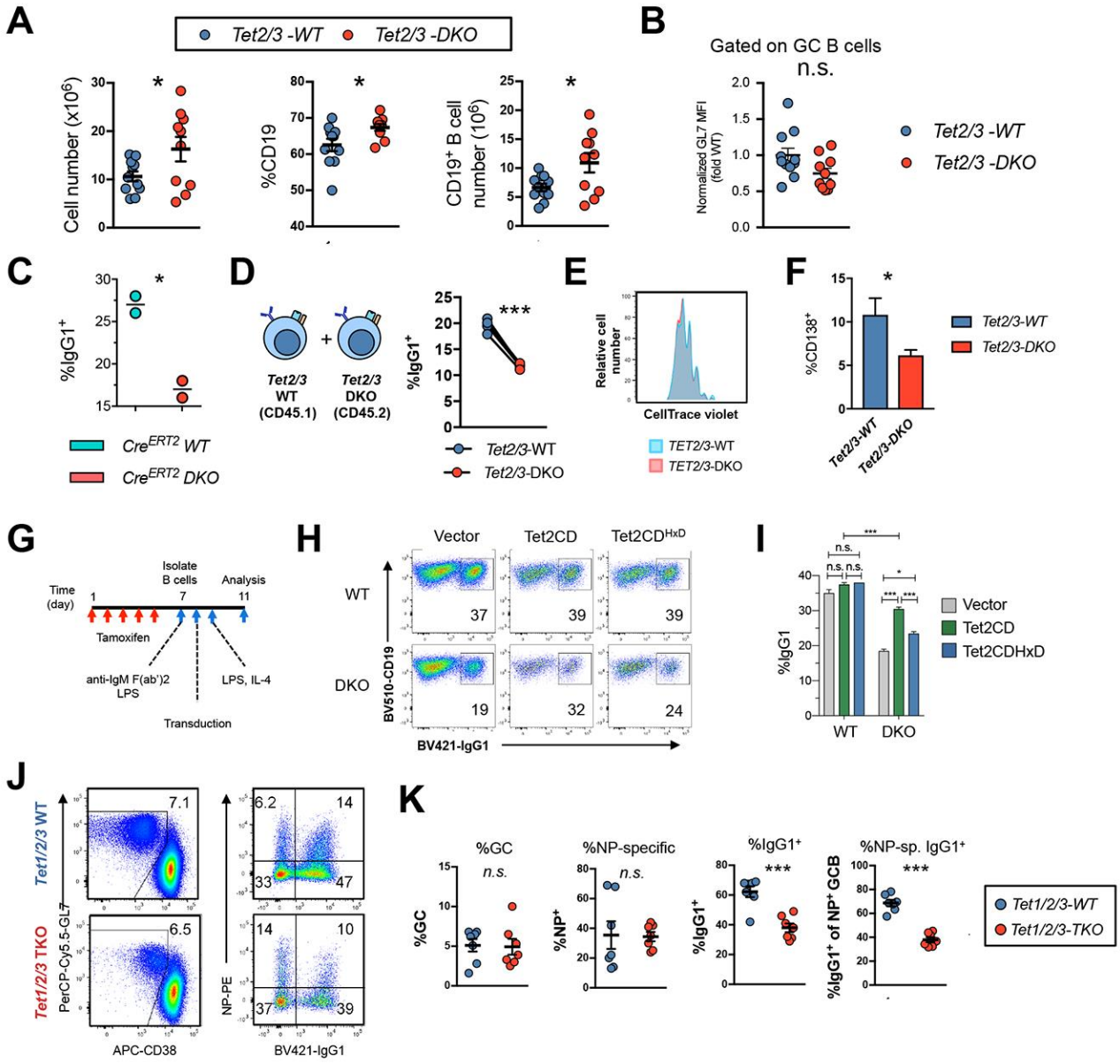

Figure S3. TET family proteins are important for B cell–intrinsic CSR.

(A) Total cell number in draining lymph nodes (left), percentage of $CD19^+$ cells (middle), and number of $CD19^+$ B cells (right) from **Fig. 3A**. *, *p<0.05*. **(B)** Level of GL7 (geometric mean of fluorescence intensity, MFI) on WT and DKO GC B cells. *, *p<0.05*. **(C) CSR defect is not caused by Cre activity.** Cre^{ERT2} Rosa26-LSL-YFP (*Cre^{ERT2} WT*) and *Cre^{ERT2} $Tet2^{fl/fl}$ $Tet3^{fl/fl}$ Rosa26-LSL-YFP* (*Cre^{ERT2}* DKO) mice were injected with tamoxifen as in **Fig. 2B**. Isolated B cells were activated with LPS and IL-4 in the presence of 4-hydroxytamoxifen for 4 days and $\%IgG1^+$ cells were analyzed (gated on live $CD19^+$ YFP^+). One representative of two experiments is shown. n=2 for each genotype. *, p<0.05. **(D-E) (D) CSR defect is cell-intrinsic.** *(Left)* *$Tet2^{+/+}$ $Tet3^{+/+}$* WT CD45.1 and *Tet2/3*-DKO CD45.2 mice were treated as in **Fig. 3A** and isolated B cells were labeled with Cell-Trace violet, 1:1 mixed, and activated with LPS and IL-4 for 4 days. *(Right)* Cells were gated based on CD45.1 and CD45.2 and the percentages of IgG1-switched cells in WT and DKO are shown. Cells from the same well are connected with lines. **(E)** Co-cultured WT and *Tet2/3*-DKO B cells showed similar proliferation profiles. Data shown are representative of two independent experiments with four technical replicates for each genotype. **(F)** Percentage of $CD138^+$ cells in the non-switched ($IgG1^-$ IgA^-) population from **Fig. 3K-3N**. **(G-I)(G)** Flow chart of experiment to assess the importance of TET catalytic activity in CSR. **(H)** Flow cytometry plots and **(I)** quantification of WT and *Tet2/3* DKO B cells transduced with empty vector (*left panels*), TET2 wild-type catalytic domain (Tet2CD, *middle panels*), and Tet2 HxD mutant catalytic domain ($Tet2CD^{HxD}$, *right panels*) shows that TET catalytic activity can partly rescue the CSR to IgG1. Data shown are representative of two independent experiments with two technical replicates. n.s., not significant. ***, *p<0.01*. *, *p<0.05*. **(J-K) Deletion of all three TET proteins (*Tet1/2/3 TKO*) results in a similar decrease in CSR as deletion of Tet2 and Tet3 (*Tet2*/3-DKO)**. **(J)** *Tet1/2/3-flox* *Cre^{ERT2}* *Rosa26-LSL-YFP* (TKO) and control *Tet1/2/3-flox Rosa26-LSL-YFP* (WT) mice were treated with tamoxifen and immunized with NP-OVA as in **Fig. 3A** and GC response and CSR were analyzed on day 7. Flow cytometry plots showed the percentage of GCB ($CD38^{lo}GL7^{hi}$) in WT and *Tet1/2/3-TKO* lymph node cells gated on total (WT) and YFP^+ (TKO) live $CD19^+$ B cells (*left panels)*. Antigen-specific (NP-PE) and class-switched cells (IgG1) were analyzed among GC B cells. **(K)** Quantification of the experiments showed in **(I)**. Data shown are aggregated results from two independent experiments. WT, n=7; TKO, n=7. Statistical significance was calculated using unpaired two-tailed *t*-test. n.s., not significant. ***, *p<0.01*. *, *p < 0.05.*

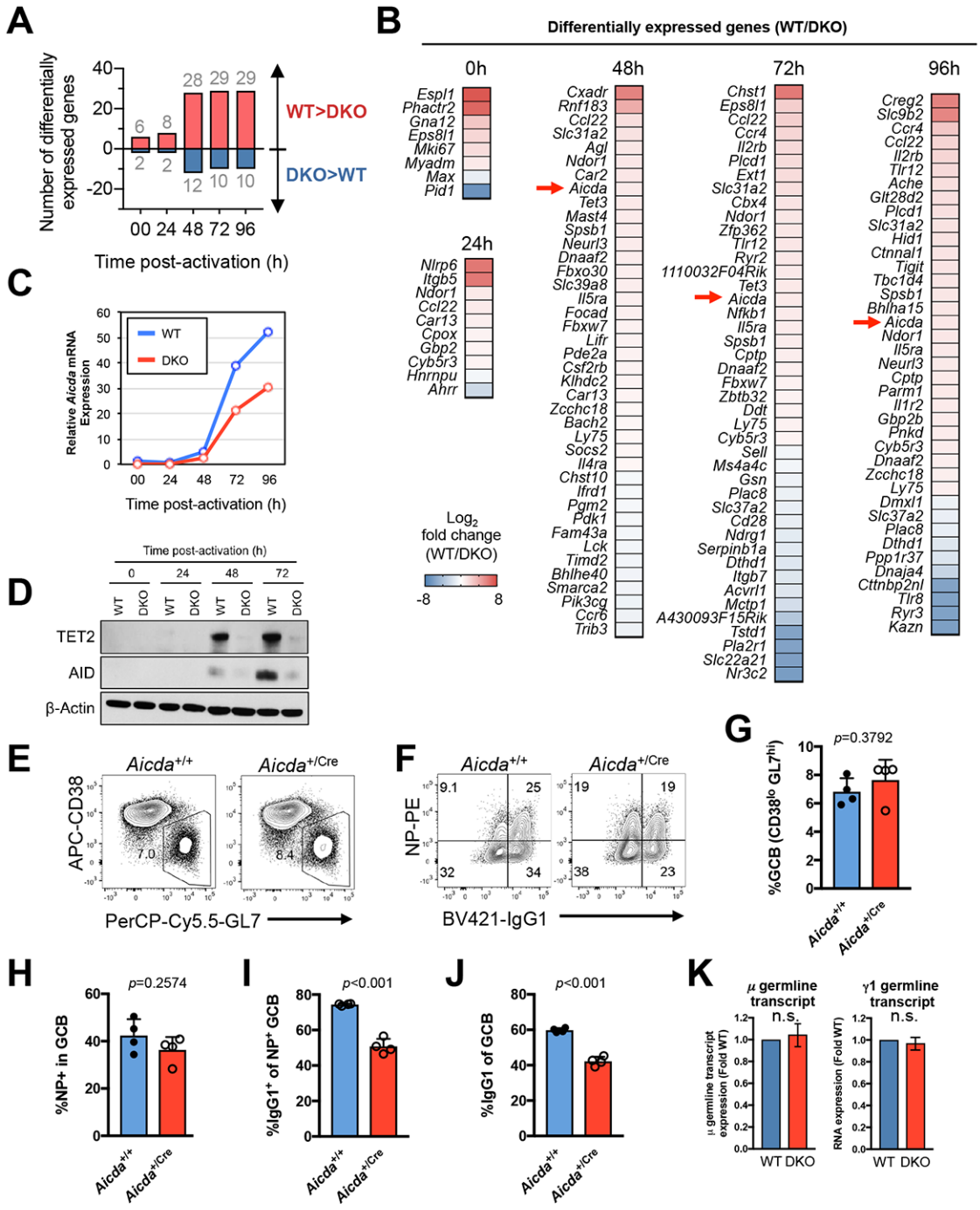

Figure S4. Decreased *Aicda* expression in *Tet2/3* DKO B cells.
WT and *Tet2/3*-DKO B cells were activated as in **Fig. 3G** and the transcriptomes were analyzed by RNA-seq (see *Materials and Methods* and **Supplementary Table S1** for details). **(A)** Number of differentially expressed genes between WT and *Tet2/3*-DKO B cells as a function of time after activation. Relatively few genes show alterations in their expression. **(B)** List of all differentially expressed genes between WT and *Tet2/3*-DKO B cells. *Aicda* (indicated by red arrows) was one of the genes expressed at significantly lower levels in DKO B cells at all time points analyzed. Color scale indicates Log_2 fold change between WT and DKO. **(C-D) TET2 and TET3 are required for potent *Aicda* expression.** *Aicda* mRNA **(C)** and protein **(D)** expression were analyzed by qRT-PCR and western blot as a function of time after activation. Results show increased AID expression with time after activation of WT B cells, and a consistent decrease in *Tet2/3* DKO relative to WT B cells. **(E-J) Haploinsufficiency of *Aicda* results in decreased CSR.** Mice with the indicated genotypes were immunized with 10 μg of NP-OVA mixed with Alum via footpad injection, and the draining lymph nodes were analyzed by flow cytometry at day 7 post-immunization. Heterozygous *Aicda-Cre* mice were used to model *Aicda* haploinsufficiency as the knocked-in Cre recombinase disrupted the endogenous *Aicda* expression. Representative flow cytometric analysis of **(E)** germinal center B cells (GCB; $CD38^{lo}$ $GL7^{hi}$) and **(F)** CSR to IgG1. **(G-J)** Statistical analyses of the populations (means and standard errors) are shown (n=4 each). Data are representative of two independent experiments. Unpaired two-tailed *t*-test was used to calculate statistical significance and the *p* values are indicated. **(K) TET2 and TET3 are not required for expression of germline transcripts.** WT and DKO B cells were activated for 4 days, and *m* and *g*1 germline transcripts were analyzed by qRT-PCR. Data were normalized to *Gapdh* and to WT level as in **Fig. 4A**. n.s., not significant.

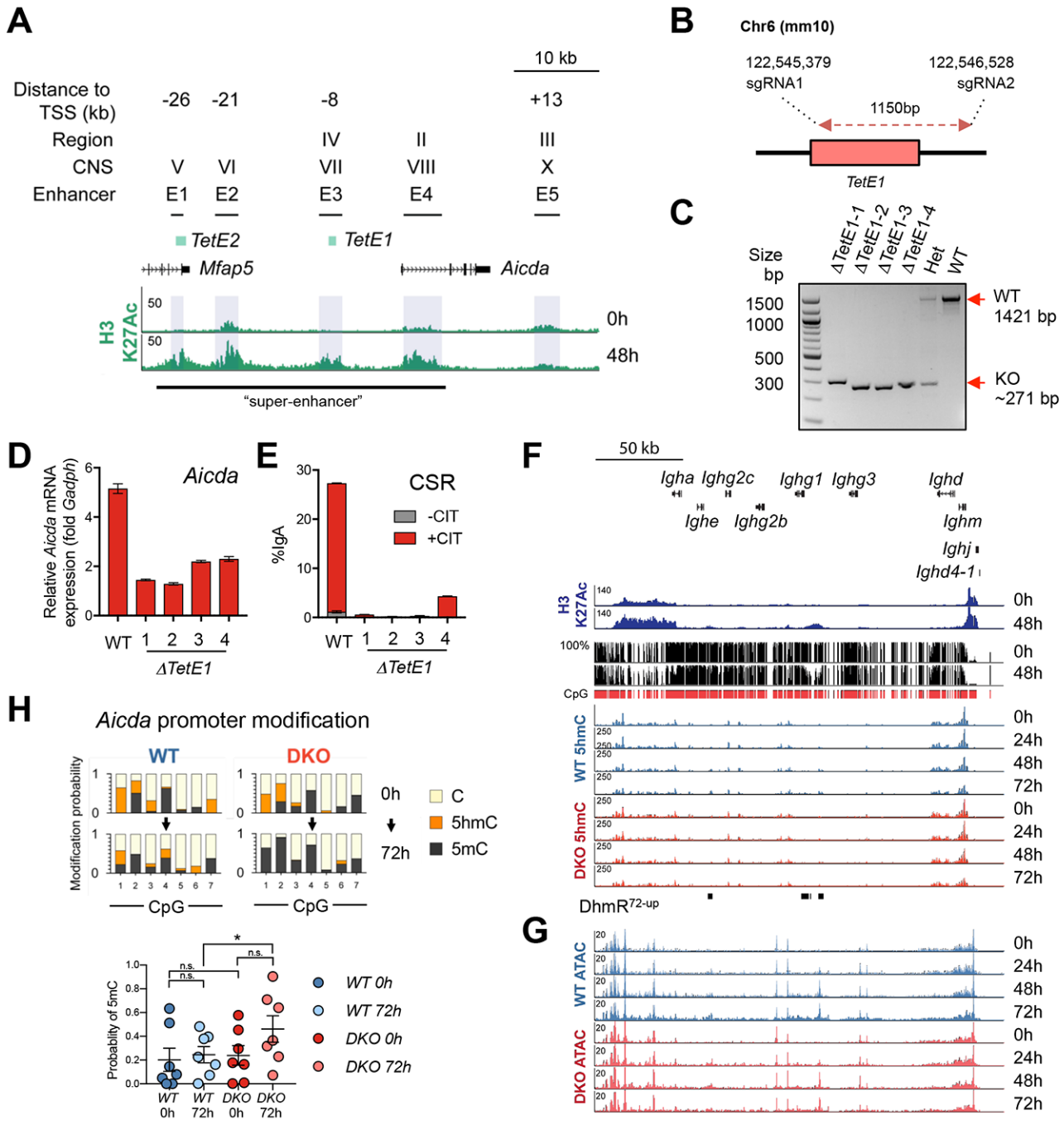

Figure S5. TET-responsive element *TetE1* regulates CSR and *Aicda* mRNA expression in the CH12 B cell line.
(A) Diagram depicts the relative position of TET-responsive elements *TetE1* and *TetE2* to previously identified *Aicda* distal and intronic enhancers. "Region" IV, II, III are from Tran et al., 2010; "CNS" V-X are from Crouch et al., 2007; "Enhancer" E1-E5 from Kieffer-Kwon et al., 2013. Note that the promoter-proximal element is not depicted. Coordinates for the shown locus are chr6:122,523,500-122,576,500 (mm10). **(B-E) *TetE1* is important for regulating *Aicda* expression and CSR. (B)** Scheme for *TetE1* deletion in CH12 cells with CRISPR. **(C)** Four clones were identified with homozygous deletion of *TetE1* as examined by PCR followed by gel electrophoresis. A clone with heterozygous deletion (Het) and a WT control are shown as controls. **(D-E)** WT and TetE1-deletion clones were stimulated with CIT (anti-CD40, IL-4, TGFβ) for two days. **(D)** *Aicda* mRNA expression and **(E)** CSR to IgA were analyzed by qRT-PCR and flow cytometry, respectively. Results show that deletion of *TetE1* decreased *Aicda* mRNA expression and abrogated CSR. **(F) 5hmC modification at the IgH locus**. Genome browser view of the IgH locus (chr12:113,211,000-113,445,000; mm10) showing H3K27Ac in unstimulated and 48h activated B cells (*top two panels*), followed by DNA methylation in unstimulated and 48h-activated B cells (*black histograms*), CpG covered in analysis (*red histograms*), and 5hmC modifications in WT and DKO as indicated. Regions with increased 5hmC modification after activation (DhmR$^{72h\text{-}up}$) are indicated by horizontal bars. **(G) Chromatin accessibility at the IgH locus.** Genome browser view of ATAC-seq data from activated WT and DKO B cells. There is no statistically significant difference in chromatin accessibility between the two genotypes. **(H) DNA modification at the *Aicda* promoter.** *Top*, CpG modifications (5hmC, 5mC, and C) *Aicda* promoter were analyzed by oxBS-seq as in **Fig. 5C**. *Bottom*, the overall CpG methylation probability was quantified. Methylation at the *Aicda* promoter at 72 h was significantly increased in DKO compared to WT. *, $p < 0.05$. n.s., not significant.

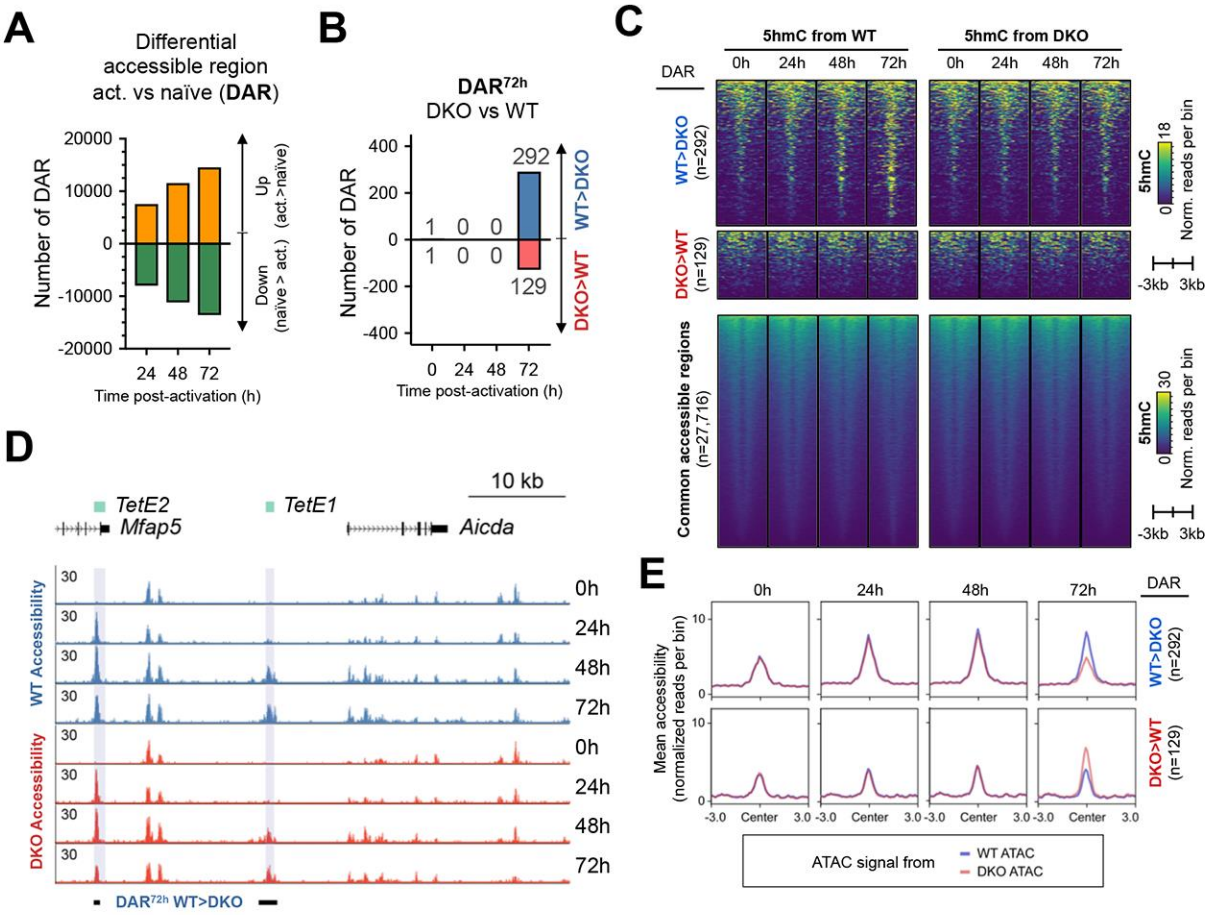

Figure S6. Tet proteins sustain enhancer accessibility.

(A) B cell activation induces global changes in chromatin accessibility. WT B cells were activated with LPS and IL-4 and chromatin accessibility was profiled by ATAC-seq at different times. Numbers indicate differentially accessible regions (DARs) between activated (act.) and naïve B cells with FDR < 0.05 and fold change above $\log_2(1.5)$ or below $\log_2(0.67)$. **(B) Loss of TET proteins results in decreased chromatin accessibility at later time points**. Numbers of DARs between WT and *Tet2/3*-DKO B cells activated for different times are shown. The difference between WT and DKO B cells was minimal at time points earlier than 72h. DARs were selected based on FDR < 0.05 and fold change above $\log_2(1.5)$ or below $\log_2(0.67)$. **(C) Tet2/3-dependent accessible regions are hydroxymethylated**. Heatmaps show the kinetics of 5hmC modification at differentially accessible regions (DARs) between WT and *Tet2/3*-DKO B cells. Regions that are more accessible in WT (WT>DKO, n=292), less accessible in WT (DKO>WT, n=129), and commonly accessible (n=27,716) are shown in the top, middle, and bottom panels, respectively. WT>DKO DARs show progressive 5hmC enrichment only in WT (*top left panels*) but not in DKO (*top right panels*) B cells, demonstrating that 5hmC modification at these regions is *Tet2/3*-dependent. The DKO>WT DARs (n=129) and common regions (n=27,716) show no apparent difference between naïve and activated B cells and between 5hmC from WT and DKO B cells. 5hmC enrichment is shown as normalized reads per 100 bp bin. **(D) TET2 and TET3 maintain chromatin accessibility at the *Aicda* Tet-responsive elements *TetE1* and *TetE2*.** Genome browser view of ATAC-seq data showing the accessibility profile at *Aicda* locus in WT (*blue, top 4 tracks*) and DKO (*red, bottom 4 tracks*) B cells. Note that *TetE1* and *TetE2* were among the DAR at 72 h (DAR^{72h} WT>DKO) as indicated at the bottom. Coordinates for the *Aicda* locus are chr6:122,523,500-122,576,500 (mm10). **(E)** Plot of mean chromatin accessibility at the DARs between WT and *Tet2/3*-DKO B cells after 72h of activation (as in **(C)** top and middle panels). *Top panels*, WT>DKO DARs (n=292); *bottom panels*, DKO>WT DARs (n=129). Y-axes indicate the mean ATAC signals (normalized ATAC-seq reads per 100 bp bin) from WT (blue line) and DKO (red line) B cells activated as indicated. The difference between WT and DKO is apparent at 72h. See also **Fig. S7**.

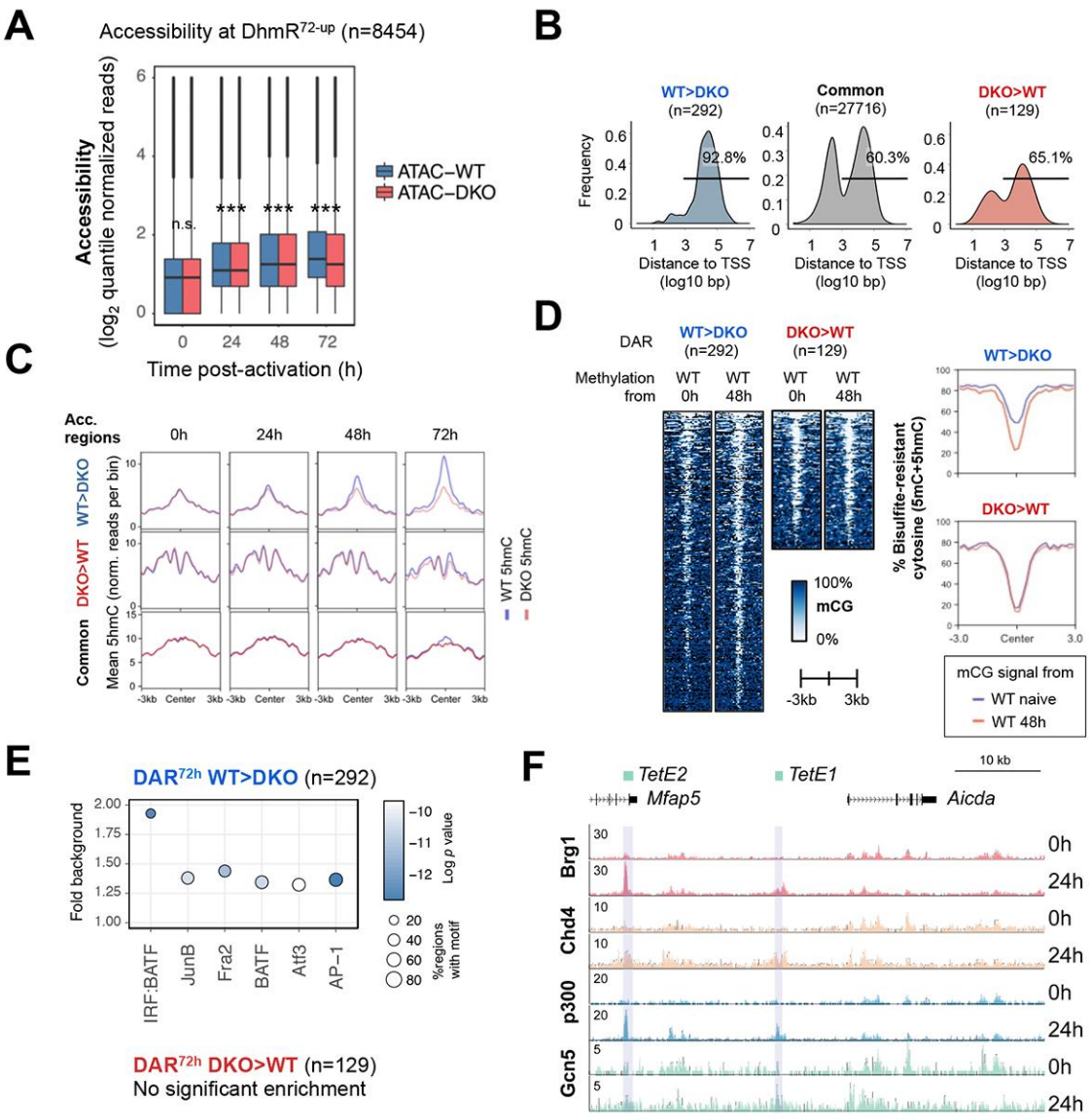

Figure S7. Analysis of TET-dependent accessible regions.

(A) Correlation between 5hmC and accessibility. Comparison of mean chromatin accessibility (ATAC-seq) between WT and DKO B cells is shown for activation-induced 5hmC-enriched regions ($DhmR^{72h\text{-}up}$, n=8454, **Fig. 1C**). Regions with increased 5hmC in WT cells after activation also show increased accessibility (blue, ATAC-WT). DKO cells show decreased accessibility, as shown for the *Aicda* locus in **Fig. S6D**. Statistical significance between WT and DKO at each time point was calculated by Kolmogorov-Smirnov test with Bonferroni correction using the family-wise error rate. n.s., not significant. ***, p adj. < 0.01. The exact adjusted *p* values are 0.06, 6.01e-05, 1.79e-07, 3.611e-11 for 0h, 24h, 48h, 72h, respectively. **(B) TET facilitates Increased accessibility at distal elements**. Histograms showing the distance of DARs (WT>DKO and DKO>WT) and commonly accessible regions (Common) from the closest TSS. Majority of the Tet-facilitated accessible regions (WT>DKO, n=292) are distal elements (>1000bp; 92.8%). **(C) TET2/3-dependent accessible regions are hydroxymethylated.** Line plots showing the kinetics of mean 5hmC modification at differentially accessible regions (DARs) between WT and *Tet2/3* DKO for the data depicted in **Fig. S6C**. Regions that are more accessible in WT (WT>DKO, n=292), less accessible in WT (DKO>WT, n=129), and commonly accessible (n=27,716) are shown in the top, middle, and bottom panels, respectively. 5hmC enrichment is shown as normalized reads per 100 bp bin. Note the Tet-dependent 5hmC modification at these DARs (*compare 72h panels*). **(D) TET-facilitated accessible regions are further demethylated after activation.** *Left,* heatmaps showing the DNA modification status (5mC+5hmC) in naïve and 48h-stimulated WT B cells at WT>DKO DARs (i.e. Tet-facilitated accessible regions; n=292) and DKO>WT DARs (n=129). *Right*, plots summarizing the data in the heatmaps; the y-axis indicates the level of bisulfite-resistant cytosine (5mC+5hmC). In WT B cells, regions that lose accessibility in *Tet2/3* DKO B cells relative to WT (WT>DKO) also show a decrease in modification (mostly 5mC) after activation. **(E)** Enrichment for consensus IRF:bZIP (IRF:BATF) and bZIP transcription factor binding motifs in the 292 Tet-facilitated accessible regions, which show increased accessibility in WT relative to *Tet2/3* DKO B cells at 72h. No significant motif enrichment was detected at DKO>WT DARs (n=129). Commonly accessible regions were used as background for the analysis. Y-axis indicates the fold enrichment versus background, circle size indicates the percentage of regions containing the respective motif, and the color indicates the significance (Log_{10} *p* value). **(F) B cell activation induces recruitment of chromatin regulators to *Aicda* distal elements.** Genome browser view of ChIP-seq data before and after B cell activation showing inducible binding of the chromatin remodelling complex components Brg1 and Chd4, and the histone acetyltransferases p300 and Gcn5, to the TET-responsive *Aicda* elements *TetE1* and *TetE2* in naïve and activated WT B cells. Scale indicates reads per 10 million. Coordinates for locus are chr6:122,523,500-122,576,500 (mm10).

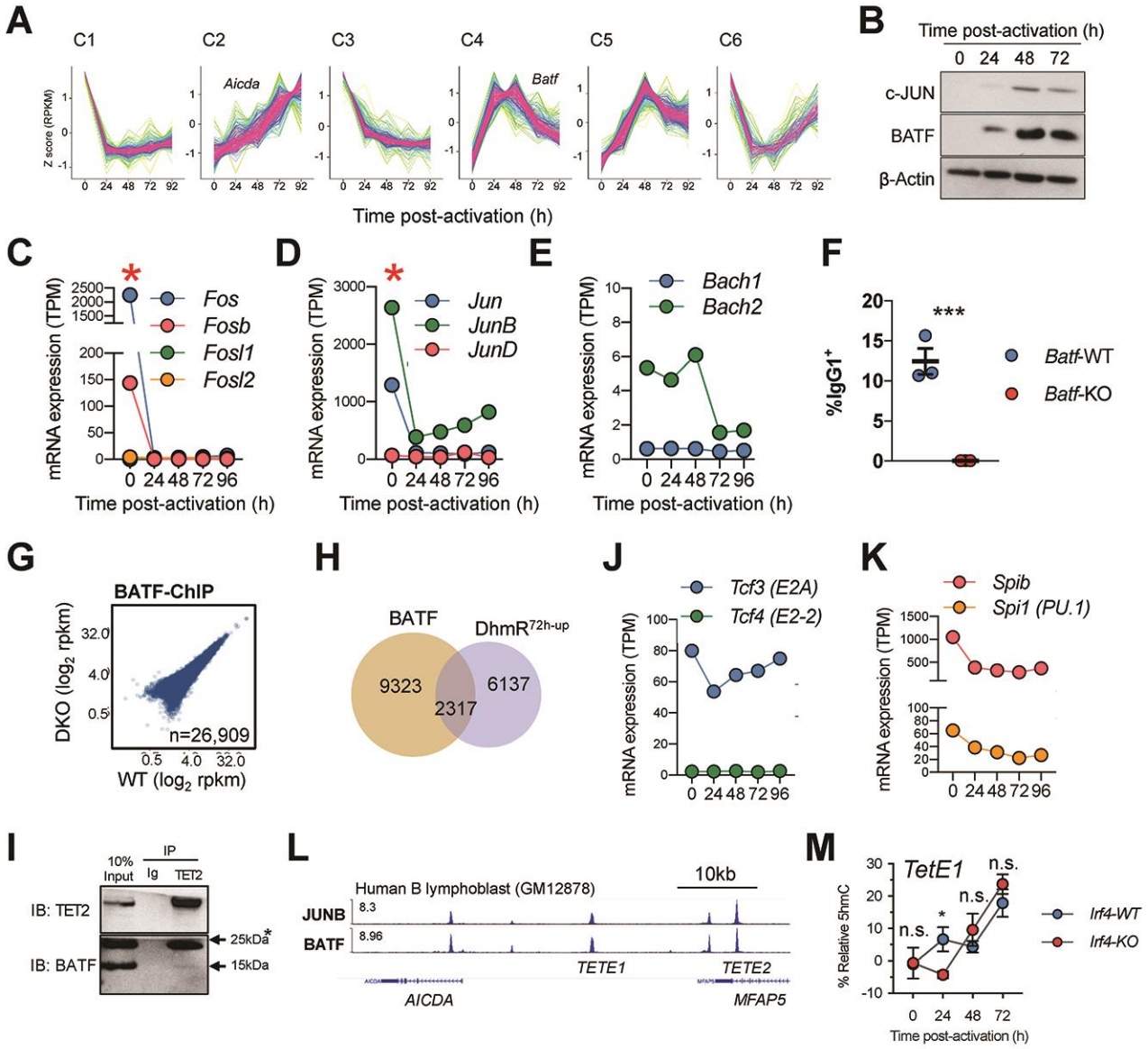

Figure S8. AP-1 proteins in activated B cells.

(A) Analysis of temporal gene expression modules (TC-seq). Gene expression was analyzed in WT cells activated for various times and genes were clustered based on their temporal expression patterns. Six clusters (C1-C6), or expression patterns, were identified. *Aicda* and *Batf* are found in C2 and C4, respectively. Y-axis indicates the Z-score calculated using RPKM. X-axis indicates the time post-activation. For details, see *Materials and Methods* and **Supplementary Table S2**. **(B-E) Expression of AP-1 proteins. (B)** Protein expression of c-JUN and BATF in unstimulated and activated B cells was analyzed by western blot. Mean mRNA expression of selected AP-1 proteins in naïve and activated B cells including *Fos* family **(C)**, *Jun* family **(D)**, and *Bach* family **(E)**. Note that the high basal expression of *Fos*, *FosB*, *Jun* and *JunB* prior to stimulation might reflect the presence of a minor population of contaminating by memory or other B cells (red asterisks)". TPM, transcript per million. **(F) BATF is required for CSR.** B cells were isolated from WT and *Batf-KO* and activated with LPS and IL-4 for 3 days. Class switch recombination to IgG1 was analyzed by flow cytometry (gated on live $CD19^{+}$). Data shown are representative of three independent experiments (n=3 each). Means and standard errors are shown. Statistical significance was calculated using unpaired two-tailed *t*-test. ***, $p<0.01$. **(G) Tet proteins are not necessary for genome-wide BATF binding**. Plots show highly similar distribution of BATF in WT and DKO 72h-activated B cells as analyzed by ChIP-seq with two independent replicates. Shown are the comparison of the BATF enrichment in WT and DKO B cells at the 26,909 regions integrated from the joined peaks from two replicates each of WT and DKO. Axes depict the $\log_2$ rpkm (read per kilobase per million) using quantile-normalized reads for each region analyzed. No region was significantly different between WT and DKO using an adjusted *p* value of 0.05. **(H) Overlap between BATF binding sites and regions with activation-induced 5hmC modification.** Venn diagram showing the number of overlapping regions between BATF peaks and $DhmR^{72h\text{-}up}$. **(I)** The interaction between BATF and TET2 was analyzed by co-immunoprecipitation using nuclear extracts from 48h-activated B cells. The pulldown was carried out in the presence of benzonase and ethidium bromide to minimize non-specific interactions via nucleic acids. The 25kDa band is non-specific as it also appears in *Batf*-KO (not shown). **(J-K) Expression of E-box and Ets family proteins.** Mean mRNA expression for E-box and Ets family proteins, from RNA-seq experiments with two independent replicates. TPM, transcripts per million. **(L) JUNB and BATF bind to *AICDA* enhancers in human B cells.** JUNB and BATF binding in human B cell lymphoblast GM12878 at the *AICDA* locus are shown (Hg38; chr12:8,598,356-8,655,770). The approximate locations for *TETE1* and *TETE2* are indicated based on sequence conservation. Data were originally from ENCODE project, processed by CistromeDB, and were viewed using WashU Epigenome Browser. **(M) *Irf4*-deficiency has no significant effect on *TetE1* hydroxymethylation**. B cells from *Cd19-Cre* (WT) and *Cd19-Cre Irf4-flox* (KO) was analyzed as in **Fig. 6E**. Two biological replicates with 3 technical replicates each. *, *p<0.05; n.s., not significant.*