PNAS
www.pnas.org

Supplementary Information for

Paradoxical association of TET loss of function with genome-wide hypomethylation

Isaac F. López-Moyado, Ageliki Tsagaratou, Hiroshi Yuita, Hyungseok Seo, Benjamin Delatte, Sven Heinz, Christopher Benner, and Anjana Rao

Anjana Rao
Email: arao@lji.org

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1903059116

Methods

External data. The external data was downloaded from Gene Expression Omnibus (GEO) and the European Nucleotide Archive (ENA). See *Table S1-S2* for details on datasets.

WGBS mapping and analysis. We employed BSMAP (v2.9) (1) to align reads from bisulfite-treated samples to the mm10 mouse reference genome allowing 4 mismatches. Reads mapping to multiple locations in the reference genome with the same mapping score were removed as well as the 5' ends bearing quality lower than 20 (mapping parameters: -n 1 -v 4 -w 2 -r 0 -q 20 -R -p 8). Single and paired-end reads were mapped as appropriately.

Duplicate reads caused by PCR amplification were removed by BSeQC (v1.0.3) (2) using default parameters. An effective genome size of 1.87e9 (as suggested in BSeQC for *Mus musculus* genome) was employed to calculate maximum coverage at the same genomic location. In addition, BSeQC was employed for removing DNA methylation artefacts introduced by end repair during adaptor ligation. For paired-end sequencing, overlapping segments of two mates of a pair were reduced to only one copy to avoid considering the same region twice during the DNA methylation quantification.

To estimate CpG DNA methylation, we employed the methratio.py tool included in BSMAP (v2.9) (1), merging DNA methylation at each CpG di-nucleotide (combining CpG methylation ratios on both DNA strands). We required each CpG to be covered by at least 5 reads (merging biological replicates) in order to be considered in the downstream analysis. Only CpGs within the autosomes were considered for the analysis (no sex chromosomes included). For the window analysis and the integration with Hi-C data, we only considered for the analysis1 kb windows with at least 3 CpGs, and 10kb windows with at least 10 CpGs.

Hi-C mapping and analysis. Reads corresponding to each extreme of a fragment were trimmed after the corresponding restriction site (e.g. MboI in the case of the NKT datasets) using HOMER (3) *homerTools trim* and independently mapped employing BWA-aln (v0.7.13) (single-end mode) (4). Reads were filtered out if they had a MAPQ score of less than 30, and only reads that were at least 25 bp were considered for the rest of the analysis. Only reads falling within the autosomes were considered for the analysis (no sex chromosomes included).

Hi-C analysis was performed using HOMER (3) and its Hi-C data analysis suite. Independently-mapped reads were paired using the makeTagDirectory command, allowing only 1 tag per bp (*-tbp 1*). Reads were filtered to remove uninformative reads (contiguous genomic fragments, self-ligation, re-ligation, and reads originating from regions of unusually high tag density) and also filtered based on the distance tor restriction sites (*-genome mm10 -removePEbg -restrictionSite GATC -both -removeSelfLigation -removeSpikes 10000 5*).

To perform the principal component analysis (PCA) of Hi-C data (A/B compartment identification), we used the tool runHiCpca.pl on the normalized interaction matrix, with the options *-res 50000 -superRes 100000 -genome mm10.* For analysis involving the Hi-C A/B compartments (e.g. integration with WGBS data), only the bins associated to the same Hi-C compartment in all biological replicates (of a given sample) were considered in the analysis.

CMS-IP and TAB-seq mapping and analysis. CMS-IP data were mapped in a similar way to WGBS. Signal per 1kb window (log_2 enrichment over input) was computed using MEDIPS (Bioconductor package) (5), using the functions MEDIPS.createSet (with the options *extend=300,*

shift=0, window_size=1000 , BSgenome="BSgenome.Mmusculus.UCSC.mm10", uniq=1e-5, paired=F for single-end data; *extend=0, shift=0, window_size=1000 , BSgenome="BSgenome.Mmusculus.UCSC.mm10", uniq=1e-5, paired=T* for paired-end data) and MEDIPS.meth (with the options *p.adj = "bonferroni", diff.method = "edgeR", minRowSum = 10, diffnorm = "tmm"*) for statistical comparisons. TAB-seq data were processed in a similar way to WGBS data.

ChIP-seq and ATAC-seq mapping and analysis. ChIP-seq and ATAC-seq data were mapped employing BWA v0.7.13 (4). Depending on the read length and sequencing type, BWA-aln was used in single or paired-end mode to map reads that were shorter than 70 bp, and reads with length >=70 bp were mapped using BWA-mem. In both cases, *Mus musculus* genome (mm10 downloaded from UCSC website) was used as reference. Reads were filtered out if they had a MAPQ score of less than 30, and only reads that were at least 25 bp were considered for the rest of the analysis. Only reads falling within the autosomes were considered for the analysis (no sex chromosomes included). For differential enrichment or occupancy analysis, the signal per 1kb genomic window was computed using MEDIPS (Bioconductor package) (5), using the functions MEDIPS.createSet (with the options *extend=300, shift=0, window_size=1000 , BSgenome="BSgenome.Mmusculus.UCSC.mm10", uniq=1e-5, paired=F* for single-end data; *extend=0, shift=0, window_size=1000 , BSgenome="BSgenome.Mmusculus.UCSC.mm10", uniq=1e-5, paired=T* for paired-end data) and MEDIPS.meth (with the options *p.adj = "bonferroni", diff.method = "edgeR", minRowSum = 10, diffnorm = "tmm"*) for statistical comparisons.

Replication timing and Lamina B data. Processed data for replication timing (6, 7) was downloaded from https://www2.replicationdomain.com/

RNA-seq mapping and transposable element (TE) analysis. Quality and adapter trimming was performed on raw RNA-seq reads using TrimGalore! 431 v0.4.5 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default parameters, retaining reads with minimal length of 25 bp. Ribosomal RNA reads were filtered out using Tagdust2. Resulting reads were aligned to mouse genome mm10 using STAR v2.5.3a (8)(Dobin et al., 2013) with alignment parameter *--outFilterMismatchNmax 4 --outFilterMultimapNmax 100 --winAnchorMultimapNmax 200*.

We employed TETranscripts (9) to quantify gene and transposable element transcript abundances. This program proportionally assigns read counts to the corresponding gene or transposable element. We used this package on mode *-multi* to be able to use ambiguously mapped reads to perform the differential expression analysis. We used the transcript annotations of the mouse genome mm10, and the repeat element annotation from UCSC RepeatMasker track of mouse genome mm10.

The DESeq2 package v1.14.1 (10) was used to normalize the raw counts and identify differentially expressed genes or transposable elements (FDR cutoff of $p<0.1$). Genes or repeat elements with less than 10 reads total were pre-filtered in all comparisons as an initial step. For total (ribodepleted) RNA-seq sample analysis, the highest expressed genes were used as control genes for size factor estimation in DESeq2. For polyA+ RNA-seq sample analysis, *p*-values of two independent experiments (same biological conditions, different library preparation methods, TruSeq and SMARTseq) were combined using the Fisher method, as implemented in the R package metaRNASeq (https://cran.r-project.org/web/packages/metaRNASeq).

Whole-genome sequencing (WGS) mapping. WGS libraries were sequenced on the Illumina Hiseq 2500 using paired-end reads at a >20X coverage per sample. Adapters and low-quality bases were trimmed before mapping, and reads with length >=70 bp were mapped to the *Mus musculus* genome (mm10 downloaded from UCSC website) using BWA-mem (4) with default options. Optical duplicate reads were removed Picard MarkDuplicates tool.

Tumor-specific variant calling. Following GATK (11) best practices for variant detection, additional pre-processing steps including recalibration of base quality scores were performed prior to variant detection. MuTect2 (12) somatic variant caller was employed to identify single-nucleotide variants (SNVs), using matched (samples Mouse A, B and C) tail information as normal (non-tumor tissue), as well as a panel of mutations observed in the recipient B6.SJL-PtprcaPep3bBoyJ mice (recipient mouse strain). In order to avoid false detection of tumor-specific SNVs (false positives), the variant calling process was repeated in a pairwise manner using the unmatched tails as normal (*e.g. Tail B and C for Mouse A*), and only SNVs detected in all three comparisons were included in the analysis. SNV filtering was performed using MuTect2 (12) default parameters. Mutational signature analysis was performed with Bioconductor's package MutationalPatterns (13). ANNOVAR (14) was used to perform functional annotation of mutations (synonymous, nonsynonymous, frameshift and nonsense mutations).

TCR repertoire analysis. The overlapping paired-end reads (250x250) were merged into a single longer read, and ClonotypeR (18) was employed to detect clonotypes in the sequence reads, extract the CDR3 sequences and quantify TCR repertoire abundances.

Mice. Mice were housed in a pathogen-free animal facility at the La Jolla Institute. They were used according to protocols approved by the Institutional Animal Care and Use Committee (IACUC). *Tet2-/-* mice were generated by crossing *CMVCre* mice to *Tet2fl/fl* mice, in which exons 8, 9 and 10 that code for the catalytic HxD domain, were floxed (flanked by LoxP sites) (15). *Tet3fl/fl* mice were generated by targeting exon 2 (16) *Tet2-/-* and *Tet3fl/fl* mice were crossed with *CD4Cre* (17) mice to generate *Tet2-/- Tet3fl/fl CD4Cre* mice (DKO mice). The *Tet2/3* DKO mice are in the C57BL/6 background. B6.SJL-PtprcaPep3bBoyJ ($CD45.1^+$) mice, C57BL/6 ($CD45.2^+$) mice mice were purchased from Jackson laboratory (B6(C)-*Cd1d1tm1.2Aben*/J). Both male and female mice were used in this study with similar findings. Invariant NKT cells were isolated from young mice (3-4 weeks old). The recipients were of the same sex as the donors. Both male and female recipients were used and similar results were obtained.

Flow Cytometry associated with NKT cell experiments. Cells were isolated from thymus, spleen, lymph nodes and bone marrow. Surface staining was performed using antibodies from Biolegend and eBioscience: CD4 (RM4-5), CD8 (53-6.7), TCRb (H57-597), B220 (RA3-6B2), CD45.1 (A20), CD45.2 (104). TCRVb2 (B20.6), TCRVb 5.1, 5.2 (MR9-4), TCRVb7 (TR310), TCRVb8.1, 8.2 (MR5-2) were purchased from BD Pharmingen. aGalCer-CD1d tetramer was obtained from the NIH Tetramer Core. Vα14*i* NKT cells were routinely defined as TCRb intermediate, B220-negative and positive for aGalCer-CD1d tetramer binding. For the pH2Ax staining the Alexa Fluor 647 anti-H2Ax-Phosphorylated (Ser139) (clone 2F3)(Biolegend) was used. Acquisition was performed in a BD LSR Fortessa (BD Biosciences) using the BD FACSDiva Software. Data analysis was performed with FlowJo (Treestar).

Isolation of Vα14*i* NKT cells. Vα14*i* NKT-cell preparations for adoptive transfer and DNA isolation experiments were performed using in case of control mice a pool of cells (isolated from thymus or spleen as indicated on each case) from C57BL/6 mice and from age- and sex-matched DKO mice.

For fluorescence-activated cell sorting (FACS), cells from wild type mice were depleted of $CD19^+$ (6D5), $TER\text{-}119^+$ (TER119), $CD8^+$ (53-6.7), $CD11c^+$ (N418), $F4/80^+$ (BM8) and $CD11b^+$ (M1/70) cells using biotinylated antibodies (Biolegend) and subsequent binding to magnetic streptavidin beads (Life Technologies). The unbound cells were incubated with 1 mg/ml Streptavidin A (Sigma Aldrich) and subsequently stained with αGalCer-loaded CD1d tetramers and anti-TCRβ, after which tetramer-binding, $TCR\beta^+$ cells were isolated using a FACSAria cell sorter (BD Biosciences). To obtain DKO cells, no depletion was performed since NKTs are massively expanded. Rather, B220-, tetramer-binding, TCRβ+ cells were isolated using a FACSAria cell sorter (BD Biosciences).

Adoptive transfer experiments. NKT-sorted cells were transferred retro-orbitrally to non-irradiated, fully immune-competent congenic (B6.SJL-PtprcaPep3bBoyJ) ($CD45.1^+$) mice.

TCR repertoire sequencing. Va14*i* NKT cells were isolated by FACS from wild type and *Tet2/3* DKO young mice or were magnetically purified by recipients of *Tet2/3 DKO* NKT cells. RNA was isolated with the E.Z.N.A. HP Total RNA kit (Omega) according to the manufacturer's instructions. cDNA was prepared using Superscript III (Invitrogen). Subsequently, PCR was performed for amplification of the gene segments with specific forward primers (*sequences shown below*) for Vb8.1 (primer MuBV8.1N), Vb8.2 (primer MuBV8.2N) and Vb8.7 (primer MuBV7) regions and a reverse primer for the b chain constant region (primer MuTCB3C). Amplicons were quantified and pooled using HS Qubit (Life Technologies). Adaptors (NEB) were ligated and libraries were amplified using Kapa HiFi (Kapa Biosystems). Amplified libraries were quantified using HS Qubit, their size was evaluated using Bioanalyzer and sequenced in an Illumina Miseq.

MuBV8.1N	GGC TGA TCC ATT ACT CAT ATG TC
MuBV8.2N	TCA TAT GGT GCT GGC AGC ACT G
MuBV7	TAC AGG GTC TCA CGG AAG AAG C
MuTCB3C	AAG CAC ACG AGG GTA GCC T

Whole-genome bisulfite sequencing (WGBS) Library preparation. $V_\alpha 14i$ NKT cells were isolated by flow cytometry and DNA was isolated using the PureLink genomic DNA mini kit (Life technologies). DNA was fragmented. 1.5 µg of the fragmented DNA was used for the library preparation and bisulfite treatment was done as described in *ref. 26*. After the bisulfite conversion the purified DNA was amplified for 4 cycles (low amplification) using Kapa HiFi $Uracil^+$ (Kapa Biosystems). 2 independent WGBS samples per genotype were evaluated.

Whole Genome Sequencing (WGS) Library preparation. Genomic DNA was isolated from purified NKT cells using the PureLink genomic DNA mini kit (Life technologies). DNA was fragmented to an average size of 400 bp using the Adaptive Focused Acoustics Covaris S2 instrument. Libraries were prepared using the TruSeq DNA PCR-Free Sample Preparation kit (Illumina) according to the manufacturer's guidelines. Libraries were purified, pooled according to the instructions of the manufacturer and sequenced in an Illumina HiSeq 2500 instrument.

Hi-C Library preparation. Between 0.6 and 1.5x10^6 NKT cells were fixed in complete medium containing 1% Formaldehyde, then quenched with 125 mM glycine and washed twice with an excess of PBS. Cells were then resuspended in lysis buffer containing 0.5% SDS and lysed at 62°C for 7 minutes. This step also allows to remove proteins that were not fixed to the chromatin. SDS was further quenched with 1% Triton-X-100 at 37°C for 15 minutes. Next, permeabilized nuclei were reacted with 100 units of MboI overnight at 37°C. After subsequent washing and

inactivation of MboI, the restriction sites were further filled in with Biotin-14-dATP and Klenow polymerase at room temperature for 40 minutes. Samples were transferred into a ligation solution containing 600 units of T4 DNA ligase. Proximity ligation was stopped by addition of 2-fold molar excess of EDTA, and samples were decrosslinked at 65°C for 16h00. DNA was further purified by proteinase K digestion and phenol/chloroform extraction. For library preparation, 800 ng of DNA was sonicated to an average of 300bp fragments length, and was used for subsequent library preparation that includes blunting of DNA, A-tailing, ligation of sequencing adapters, and amplification of library.

Total (ribodepleted) RNA-seq Library preparation. 10 million cells were sorted and then whole RNA was isolated using the RNeasy Plus Mini Kit (Qiagen). Ribo-zero RNA-seq libraries were prepared using the TruSeq Stranded Total RNA Library Prep Gold kit (Illumina) with minor modifications. The starting RNA was 800 ng. Ribosomal RNAs were depleted using magnetic beads. Next, RNA was fragmented, and cDNA was synthesized using Superscript II (Invitrogen). After A-tailing and adaptor ligation, libraries were generated by amplifying the cDNA for 12 cycles.

Statistical Analysis. For mouse experiments, Mantel-Cox test and Gehan-Brenslow-Wilcoxon test were applied as indicated and the p values are shown for each figure. Statistical evaluations were performed using the unpaired t test. Data are mean $\pm$ SEM. Asterisks indicate statistically significant differences: ****P< 0.0001, ***P< 0.001, **P< 0.01, *P<0.05. If not otherwise indicated the p value was not statistically significant (p>0.05). In the graphs each dot represents a mouse. For all the experiments we used sufficient number of mice to ensure adequate power for our conclusions. Mice from different litters and of different sex were evaluated. In addition, we ensured that a minimum of 2 independent experiments was performed in each case. No blinding was applied in this study. For the two-sample Kolmogorov-Smirnov test related to methylation analysis, the *D* statistic and pvalues were calculated using the *ks.test* function as implemented in R. In all tests, the alternative hypothesis is that CDF of WT lies below that of *TET*.

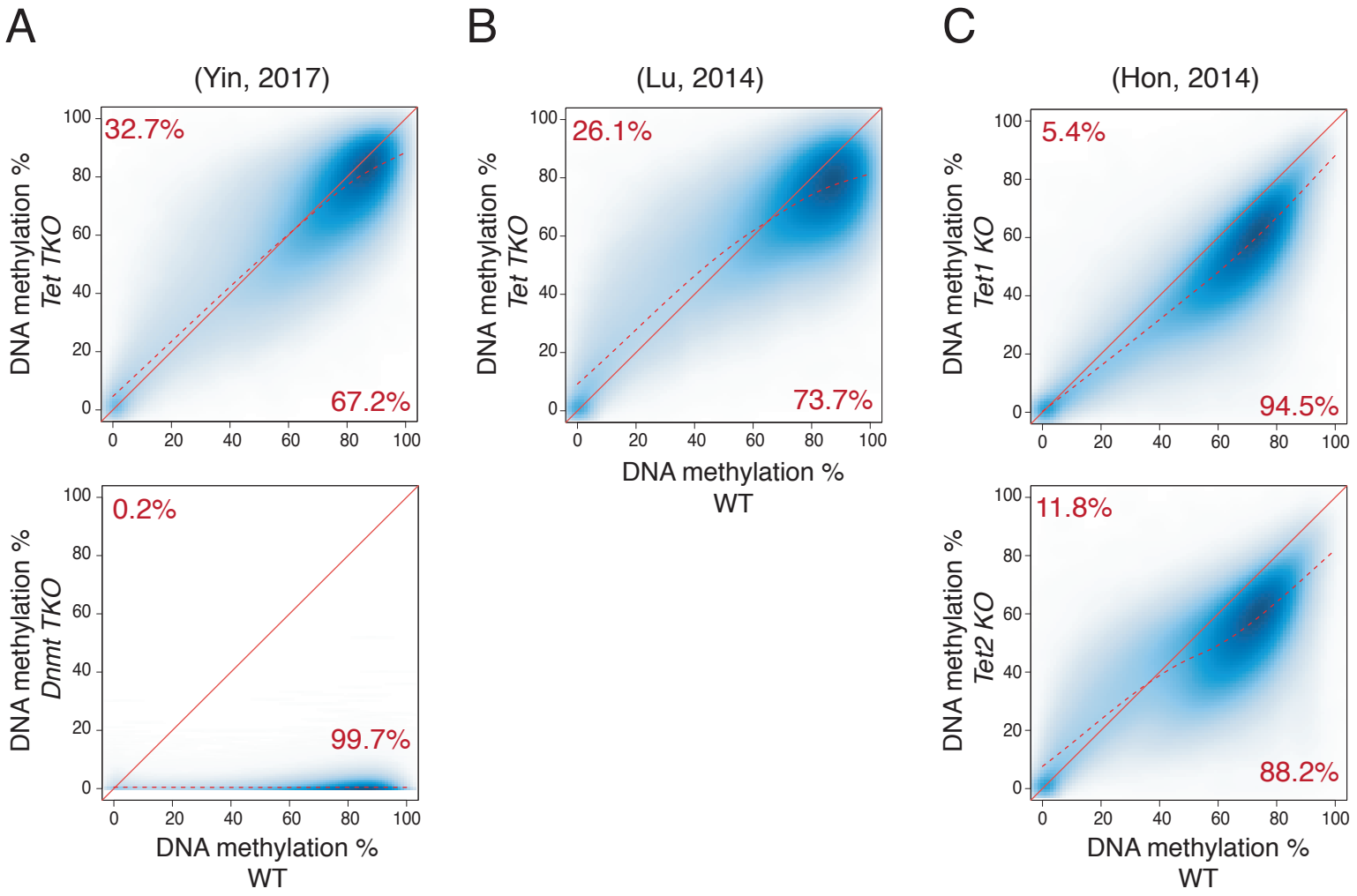


Fig. S1. B compartment hypomethylation in *TET*-mutant mESC. (*A*), Smoothed scatterplot of the average DNA methylation values within 1 kb windows across the genome, comparing WT cytosine modification values (x-axis) to the ones in *Tet TKO* (*top*), and *Dnmt TKO* (*bottom*) (y-axis). LOESS regression (dashed line) is displayed for each panel. (*B*), Smoothed scatterplot of the average DNA methylation values within 1 kb windows across the genome, comparing WT cytosine modification values (x-axis) to the ones in *Tet TKO* (y-axis). LOESS regression (dashed line) is displayed for each panel. (*C*), Smoothed scatterplot of the average DNA methylation within 1 kb windows across the genome, comparing cytosine modification values between WT (x-axis) and *Tet1 KO* or *Tet2 KO* (y-axis) mESC. LOESS regression (dashed line) is displayed for each panel. *Tet1 KO* mESC (*top*) show global loss of methylation whereas *Tet2 KO* mESC (*bottom*) show hypermethylation in regions with low-intermediate methylation in WT mESC and hypomethylation at regions with high methylation in WT mESC.

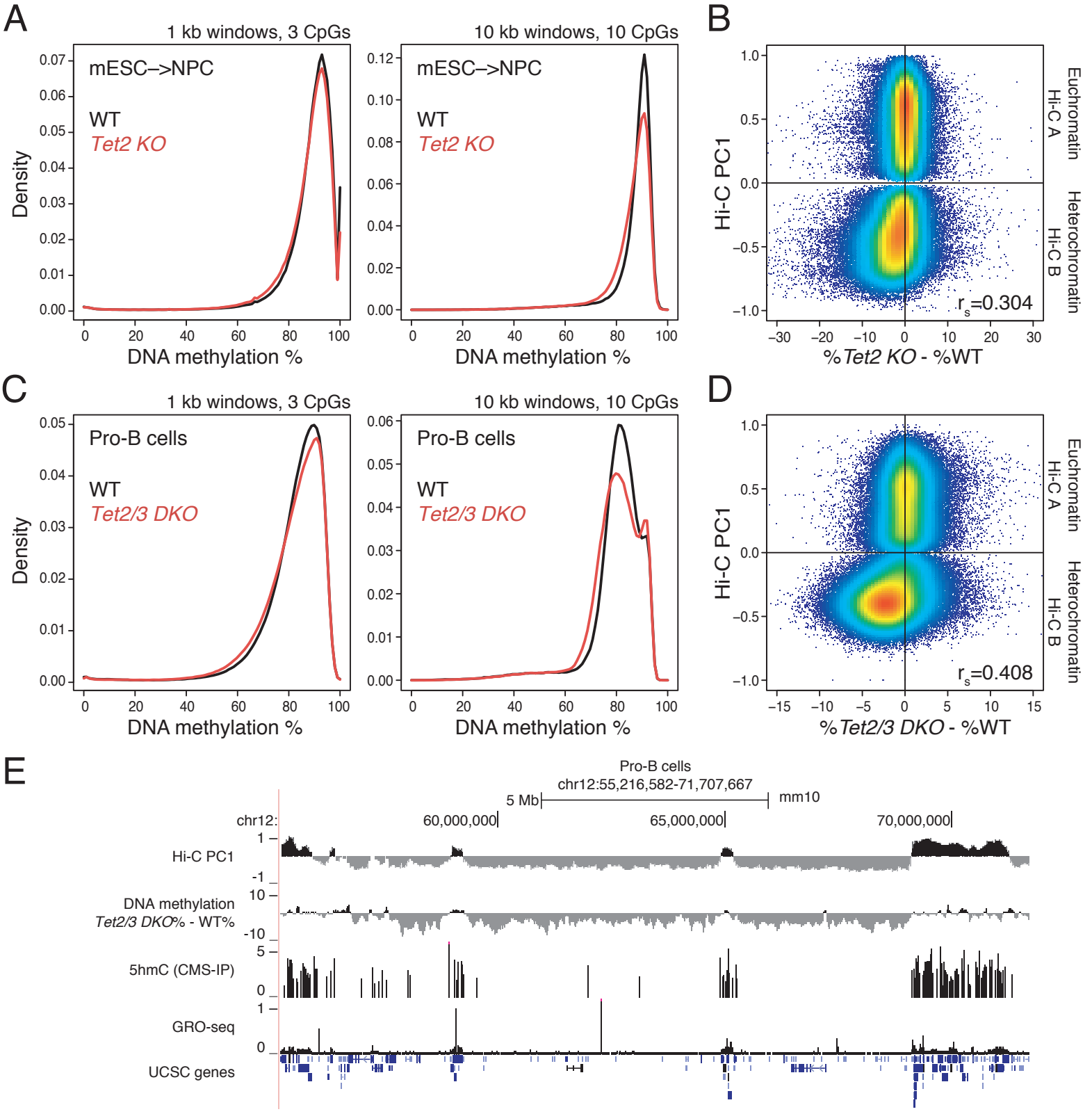
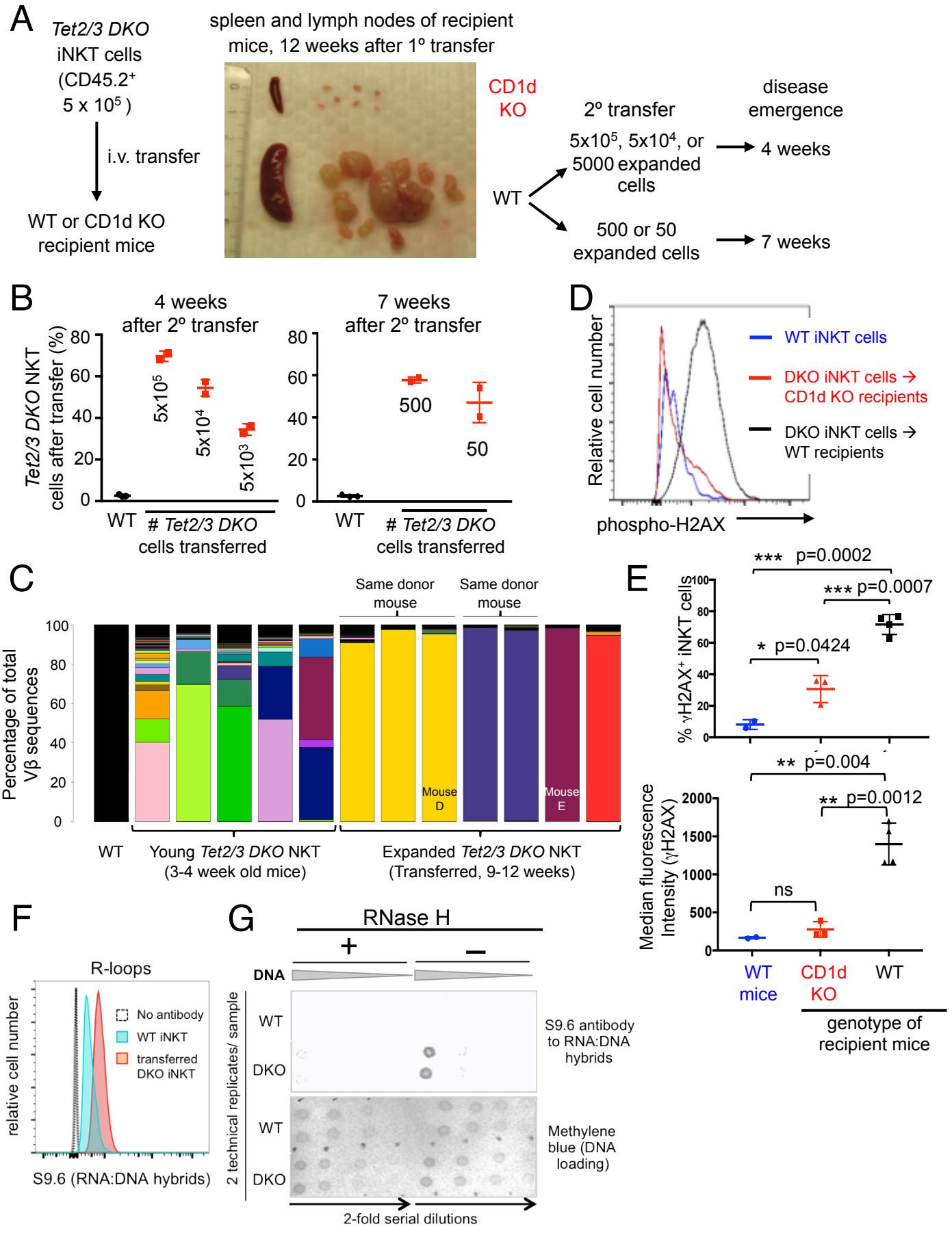


Fig. S2. B compartment hypomethylation in *TET*-mutant NPC and pro-B cells. (*A*), Distribution of average DNA methylation values within 1 kb and 10 kb windows across the genome in wild type (WT) and *Tet2 KO* mESC differentiated to neural precursor cells (NPC) at day 3 post-differentiation. (*B*), Correlation between DNA methylation changes (difference in cytosine modification percentage, *Tet2 KO* minus WT) and A/B compartments (Hi-C PC1 values) in mESC differentiated to NPCs. The Spearman correlation coefficient is shown (r_s value). (*C*), Distribution of average DNA methylation values within 1 kb and 10 kb windows across the genome, in wild type (WT) and *Tet2/3 DKO* pro-B cells. (*D*), Correlation between DNA methylation changes (difference in cytosine modification percentage, *Tet2/3 DKO* minus WT) and A/B compartments (Hi-C PC1 values) in mouse pro-B cells. The Spearman correlation coefficient is shown (r_s value). (*E*), Genome tracks showing a correspondence between heterochromatic Hi-C B compartment and large hypomethylated domains in pro-B cells (shown as subtraction of DNA methylation percentage, *Tet2/3 DKO* minus WT). RNA transcription (GRO-seq track) and 5hmC distribution (CMS-IP track) in WT pro-B cells are shown for reference.



(*Legend on next page*)

Fig. S3. Expansion of *Tet2/3 DKO* NKT cells is accompanied by increased clonality, and accumulation of DNA double-strand breaks and R-loops. (*A*), Left, experimental workflow. Middle, the picture depicts splenomegaly and enlarged lymph nodes in wild type (WT) but not *CD1d KO* recipients of *Tet2/3 DKO* NKT cells. Right, times of disease emergence (see (*B*)). (*B*), Percentage of WT or expanded *Tet2/3 DKO* NKT cells in splenocytes of congenic WT recipient mice, injected with the indicated numbers of *Tet2/3 DKO* NKT cells. 2 mice were used per condition. Mice injected with as few as 50 *Tet2/3 DKO* cells develop NKT cell lymphoma. (*C*), Evaluation of TCRβ chain CDR3 variable region sequences in DNA from NKT cells. One WT mouse, five 3-4 week old *Tet2/3 DKO* mice, and seven WT recipients of 0.5 million *Tet2/3 DKO* NKT cells were evaluated. Each color represents a single TCR Vβ sequence. Mice that received NKT cells from a single *Tet2/3 DKO* donor showed expansion of the same NKT cell clone. Asterisks indicate independent mice Mouse D and Mouse E for which SNV data are shown in Fig. 4B. (*D*), Evaluation of phospho-H2AX staining as a marker for DNA DSBs in WT and *Tet2/3 DKO* NKT cells. NKT cells (αGalCer-CD1d+, TCRβ+) were isolated from healthy WT mice or after transfer of *Tet2/3 DKO* NKT cells to non-irradiated WT or *CD1d KO* recipients as outlined in (*A*). A representative flow cytometric analysis is shown. (*E*), (Top panel) Percentage of phospho-H2AX+ NKT cells isolated from WT mice or from *Tet2/3 DKO* NKT cells transferred to and recovered from non-irradiated WT or *CD1d KO* recipients as outlined in (*A*). (Bottom panel) Median fluorescence intensity of phospho-H2AX staining in NKT cells isolated from WT or from *Tet2/3 DKO* mice after transfer to and recovery from WT or *CD1d KO* recipients as outlined in (*A*). Data are mean ± SEM, n=2 (WT NKT cells), n=3 (NKT cells from *CD1d KO* recipients) and n=4 (NKT cells from WT recipients). ns, not significant. At least 2 independent experiments were performed per condition. (*F*), Flow cytometric analysis evaluating R-loops in NKT cells isolated from healthy WT mice (blue histogram) or from *Tet2/3 DKO* NKT cells transferred and expanded in nonirradiated immunocompetent WT recipients (red histogram). The S9.6 antibody recognizes RNA:DNA hybrids. (*G*), Dot blot of genomic DNA from *Tet2/3 DKO* NKT cells transferred to and expanded in non-irradiated WT recipients shows increased R-loop formation compared to WT NKT cells (top panel, right). Specificity for RNA:DNA hybrids was confirmed by RNase H treatment of genomic DNA prior to spotting, which results in elimination of the signal (top panel, left). Equivalent DNA loading was confirmed by methylene blue staining (bottom panel).

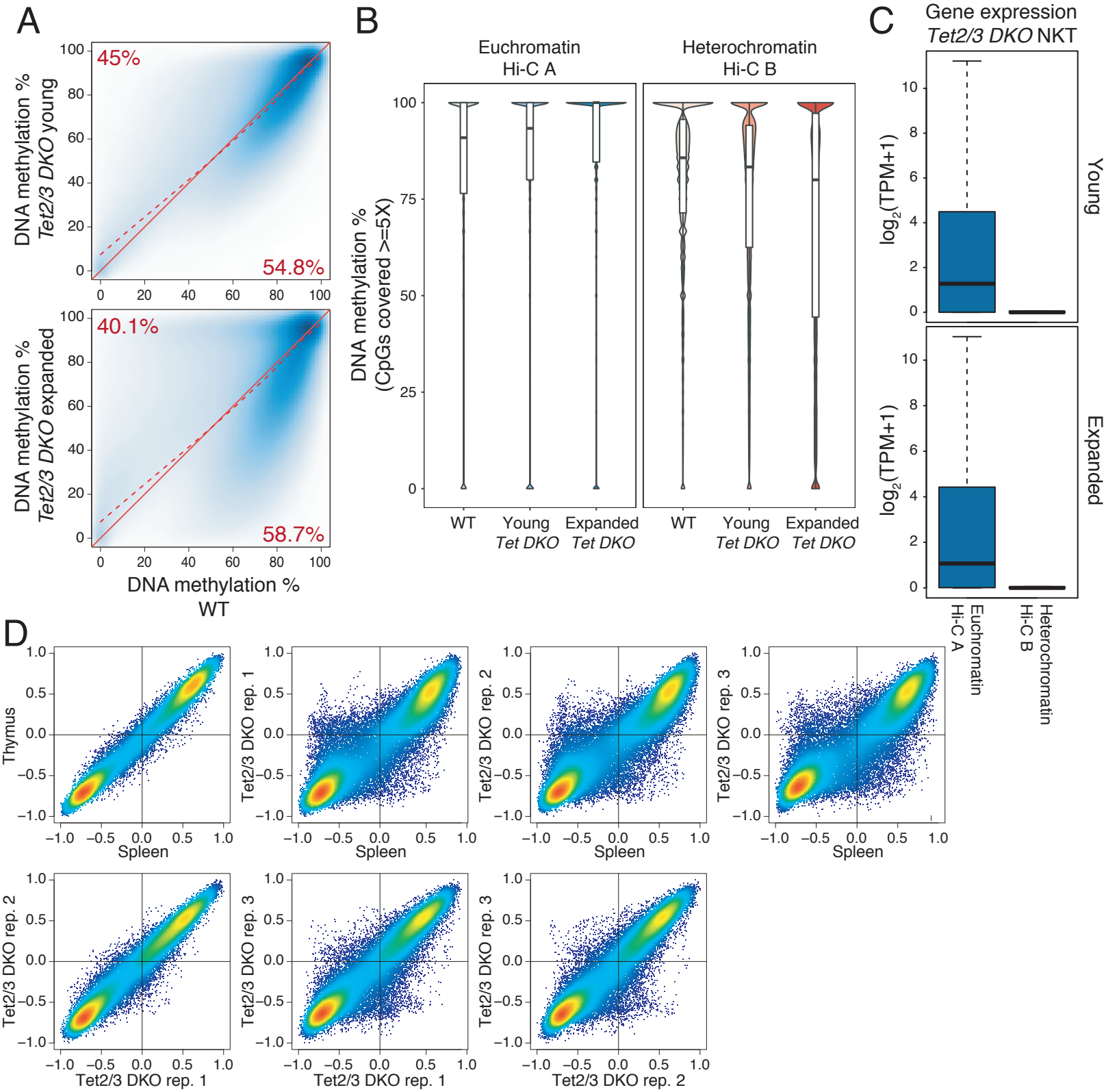
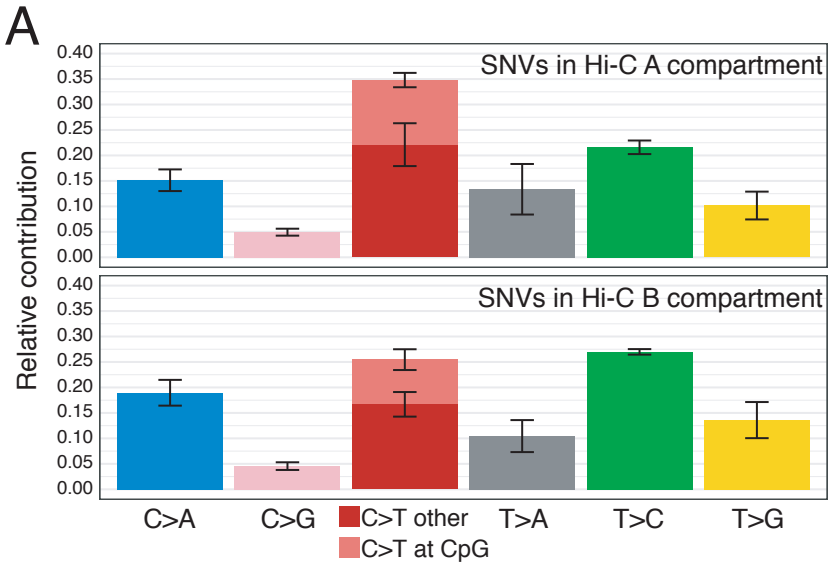
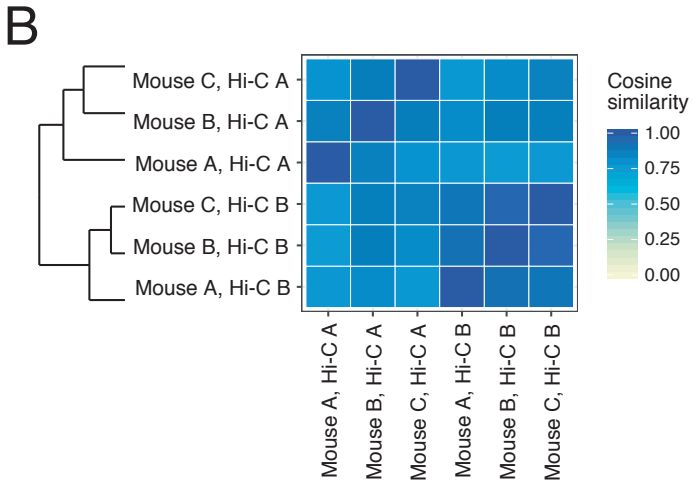
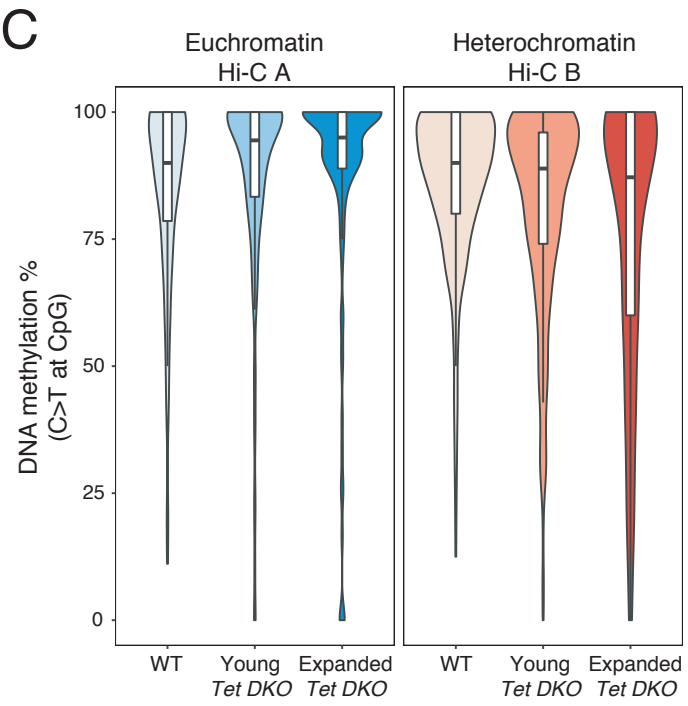
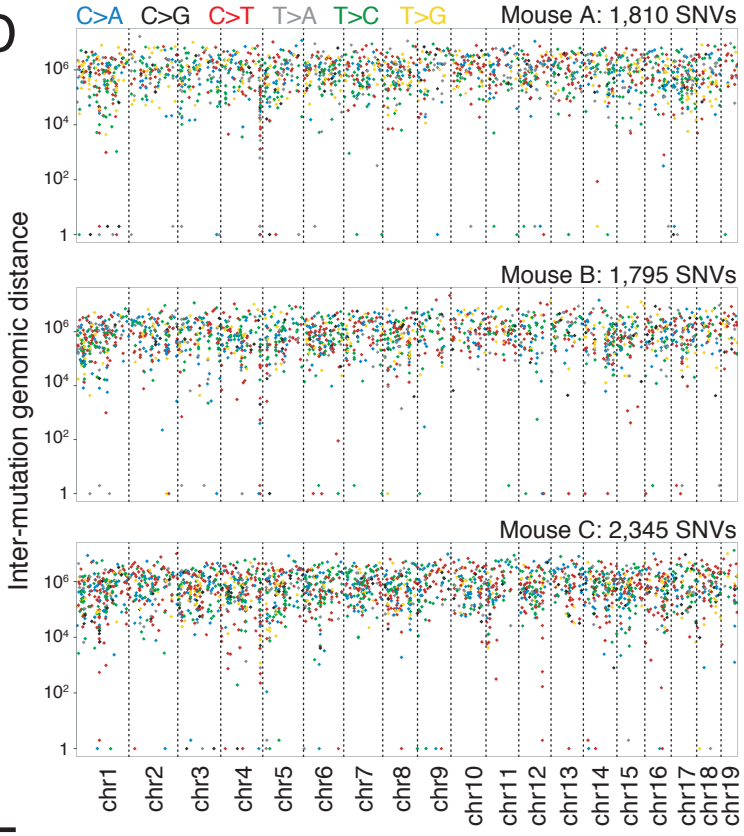


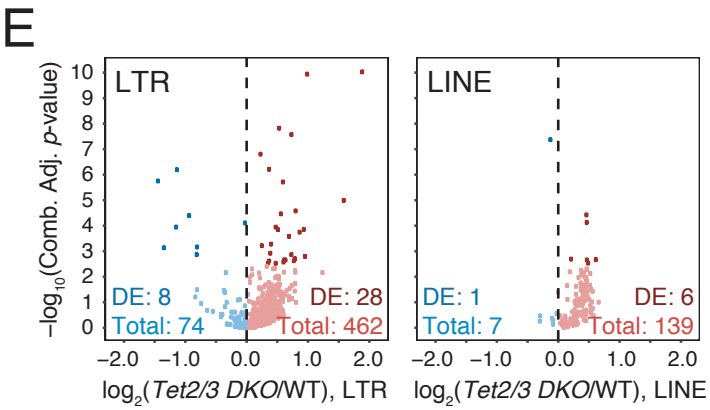
Fig. S4. *Tet2/3*-deficient NKT cell lymphoma displays progressive hypomethylation in the heterochromatic Hi-C B compartment. (*A*), Smoothed scatterplot of the average DNA methylation within 1 kb windows across the genome, comparing WT (x-axis) to young (left panel) or transferred and expanded (right panel) *Tet2/3 DKO* NKT cells (y-axis). LOESS regression (dashed line) is displayed for each panel. (B), Distribution of DNA methylation (WGBS signal) at cytosines within the CpG context covered by at least 5 WGBS reads. DNA methylation values are shown for WT, *Tet2/3 DKO* young, and expanded *Tet2/3 DKO* NKT cells. (C), Comparison of gene expression levels (RNA-seq, log2(TPM+1)) in the euchromatic Hi-C A and heterochromatic Hi-C B compartments in young (left panel) and expanded (right panel) *Tet2/3 DKO* NKT cells. (D), Pairwise comparison of Hi-C PC1 values between independent samples. Note the good correlation between the WT thymus and spleen NKT cell samples and between *Tet2/3 DKO* replicates 1 and 2 which were expanded from the same donor mouse, compared to the slightly greater differences between *Tet2/3 DKO* replicates 1 and 2 and replicate 3 which was from a different donor mouse. Similarly, there are slight differences in the Hi-C compartment between NKT cells from WT spleen and all three *Tet2/3 DKO* NKT cells taken from the spleen.











F

Sample	Coverage	Tail control
Mouse A	47.69	Matched tail
Mouse B	37.89	Matched tail
Mouse C	25.48	Matched tail
Mouse D	20.79	Unmatched
Mouse E	18.33	Unmatched

(*Legend on next page*)

Fig. S5. Mutational spectrum of transferred and expanded *Tet2/3 DKO* T cells. (*A*), Mutational spectrum associated with euchromatic Hi-C A and heterochromatic Hi-C B compartments in *Tet2/3 DKO* expanded NKT cells. (*B*), Cosine similarity between mutational profiles obtained from three independent *Tet2/3 DKO* expanded T cells WGS samples (matched tail samples, Mouse A-C), separating the mutations by their location within Hi-C compartments (A vs B) in *Tet2/3 DKO* NKT cells. Notice how mutational profiles cluster by Hi-C compartment and not by sample of origin. (*C*), DNA methylation at cytosines within the CpG context with C>T substitution type. DNA methylation values shown for WT, *Tet2/3 DKO* young, and *Tet2/3 DKO* expanded T cells. (*D*), Rainfall plots representing the inter-mutational genomic distance (y-axis) for all single nucleotide variants (SNV) (x-axis) encountered in samples Mouse A-C. Substitutions are color-coded as indicated in the top of the figure. While mutations cluster at certain regions, each mice exhibits a unique spectrum of mutations. (*E*), Distribution of the changes in expression ($\log_2$ fold change difference) of LTR and LINE transposable elements in *Tet2/3 DKO* young NKT cells compared to WT, obtained from analysis of polyA+ RNA-seq data. *P*-values of two independent experiments (same biological conditions, different library preparation methods, TruSeq and SMART-seq) were combined using the Fisher method. (*F*), Table with coverage values of WGS samples.

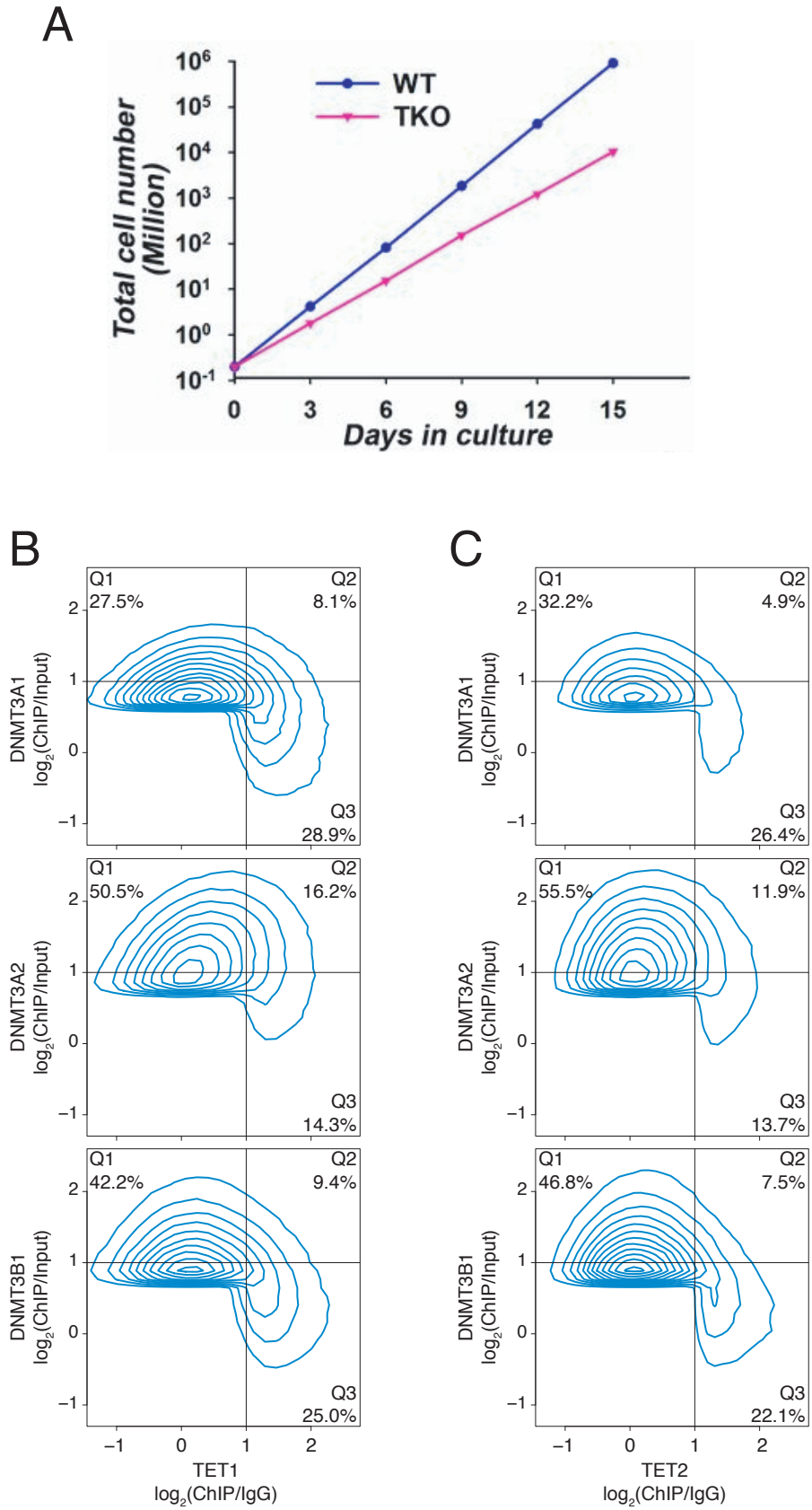


Fig. S6. *TET TKO* mESC proliferate more slowly than their WT counterparts, and genome-wide distribution of DNMT3 enzymes in relation to TET1 and TET2 in mESC. (A), Growth curves of WT and *Tet1/2/3 TKO* mESCs. Cells were split every 3 days, and cells were counted. Reprinted from ref. 16. (B-C), Comparison of the localization of DNMT3 proteins versus TET1 (B) and TET2 (C) in WT mESC. ChIP-seq enrichment (log_2 fold change) was calculated for 1 kb windows. A different dataset (51, 52) was used from that of Fig. 7.

Supplementary table S1. Cell types/Methylomes analyzed in this study

Cell type	Reference	Laboratory	Genotype	Tet1	Tet2	Tet3	Other
ESC	Yin et al, Science, 2017	Jussi Taipale	WT	WT	WT	WT	
ESC	Yin et al, Science, 2017	Jussi Taipale	Tet1/2/3 TKO	KO	KO	KO	
ESC	Yin et al, Science, 2017	Jussi Taipale	Dnmt1/3a/3b TKO	WT	WT	WT	DNMT TKO
ESC	Lu et al, Genome Research, 2014	Yi Zhang	WT	WT	WT	WT	
ESC	Lu et al, Genome Research, 2014	Yi Zhang	Tet1/2/3 TKO	KO	KO	KO	
ESC	Hon et al, Molecular Cell, 2014	Bing Ren	WT	WT	WT	WT	
ESC	Hon et al, Molecular Cell, 2014	Bing Ren	Tet1 KO	KO	WT	WT	
ESC	Hon et al, Molecular Cell, 2014	Bing Ren	Tet2 KO	WT	KO	WT	
ESC->NPC	Hon et al, Molecular Cell, 2014	Bing Ren	WT	WT	WT	WT	
ESC->NPC	Hon et al, Molecular Cell, 2014	Bing Ren	Tet2 KO	WT	KO	WT	
NKT	Tsagaratou et al, Nature Immunology, 2017	Anjana Rao	WT	WT	WT	WT	
NKT	Tsagaratou et al, Nature Immunology, 2017	Anjana Rao	Tet2/3 DKO (young)	WT	KO	KO	
NKT	This study	Anjana Rao	Tet2/3 DKO (expanded)	WT	KO	KO	
proB	Lio et al., eLife, 2016	Anjana Rao	WT	WT	WT	WT	
proB	Lio et al., eLife, 2016	Anjana Rao	Tet2/3 DKO	WT	KO	KO	
HSC	Jeong et al, Nature Genetics, 2029	Margaret Goodell	WT	WT	WT	WT	
HSC	Zhang et al, Nature Genetics, 2016	Margaret Goodell	Tet2 KO	WT	KO	WT	
HSC	Jeong et al, Nature Genetics, 2029	Margaret Goodell	Dnmt3a KO	WT	WT	WT	Dnmt3a KO
HSC	Zhang et al, Nature Genetics, 2016	Margaret Goodell	Dnmt3a/Tet2 DKO	WT	KO	WT	Dnmt3a KO
ESC	Baubec et al, Nature, 2015	Dirk Schübeler	Dnmt TKO+Dnmt3a2	WT	WT	WT	Dnmt TKO+Dnmt3a2
ESC	Baubec et al, Nature, 2015	Dirk Schübeler	Dnmt TKO+Dnmt3b1	WT	WT	WT	Dnmt TKO+Dnmt3b1
ESC	Manzo et al, EMBO J, 2017	Tuncay Baubec	Dnmt TKO+Dnmt3a1	WT	WT	WT	Dnmt TKO+Dnmt3a1

Supplementary table S2: Publicly available datasets used in this study

Experiment type	Experiment	Accession Number	Description	Reference	Laboratory
WGBS	WGBS_ESC	PRJEB9797	WGBS, WT, Tet TKO and Dnmt TKO mESC	Yin et al., 2017	Jussi Taipale
WGBS	WGBS_ESC	GSE56986	WGBS, WT and Tet TKO mESC	Lu et al., 2014	Yi Zhang
WGBS	WGBS_ESC	GSE48519	WGBS, WT, Tet1 KO and Tet2 KO mESC	Hon et al., 2014	Bing Ren
WGBS	WGBS_NKT	GSE66834	WGBS, WT and young Tet2/3 DKO NKT cells	Tsagaratou et al., 2017	Anjana Rao
WGBS	WGBS_HSC	GSE49714	WGBS, WT and Dnmt3a HSC	Jeong et al., 2014	Margaret Goodell
WGBS	WGBS_HSC	GSE72148	WGBS, Tet2 KO and Tet2/Dnmt3a KO HSC	Zhang et al., 2016	Margaret Goodell
WGBS	WGBS_NPC	GSE48519	WGBS, WT and Tet2 KO, mESC->NPC	Hon et al., 2014	Bing Ren
WGBS	WGBS_proB	SRP076056	WGBS, WT and Tet2/3 DKO proB cells	Lio et al., 2017	Anjana Rao
WGBS	WGBS_ESC_DNMT_TKO	GSE57411	WGBS, DNMT TKO ESC, TKO+DNMT3A2, TKO+DNMT3B1	Baubec et al., 2015	Dirk Schübeler
WGBS	WGBS_ESC_DNMT_TKO	GSE96087	WGBS, DNMT TKO ESC, TKO+DNMT3A1	Manzo et al., 2017	Tuncay Baubec
Hi-C	HiC_ESC	GSE96107	Hi-C, ESC	Bonev et al., 2017	Giacomo Cavalli
Hi-C	HiC_HSC	GSE79422	HiC, HSC	Hu et al., 2018	Keji Zhao
Hi-C	HiC_NPC	GSE96107	Hi-C, NPC	Bonev et al., 2017	Giacomo Cavalli
Hi-C	HiC_proB	GSE40173	Hi-C pro-B, method: FA and EGS	Lin et al., 2012	Cornelius Murre/Chris Glass
RNA-seq	RNA_ESC	GSE48519	RNAseq, WT mESC	Hon et al., 2014	Bing Ren
RNA-seq	RNA_HSC	GSE49714	RNAseq, WT HSC	Jeong et al., 2014	Margaret Goodell
RNA-seq	RNA_NKT	GSE66834	RNAseq, SMARTseq, WT and young Tet2/3 DKO NKT cells	Tsagaratou et al., 2017	Anjana Rao
RNA-seq	RNA_NKT	GSE66834	RNAseq, Truseq, WT and young Tet2/3 DKO NKT cells	Tsagaratou et al., 2017	Anjana Rao
GRO-seq	GRO_proB	GSE40173	GRO-seq pro-B	Lin et al., 2012	Cornelius Murre/Chris Glass
CMS-IP-seq	CMS_HSC	GSE49714	CMS-IP, WT HSC	Jeong et al., 2014	Margaret Goodell
CMS-IP-seq	CMS_NKT	GSE66834	CMS-IP, WT NKT cells	Tsagaratou et al., 2017	Anjana Rao
CMS-IP-seq	CMS_proB	SRP076056	CMS-IP, WT proB cells	Lio et al., 2017	Anjana Rao
TAB-seq	TAB_ESC	GSE48519	TABseq, WT mESC	Hon et al., 2014	Bing Ren
ATAC-seq	ATAC_NKT	GSE66834	ATACseq, WT and young Tet2/3 DKO NKT cells	Tsagaratou et al., 2017	Anjana Rao
ChIP-seq	ChIP_H3K27me3_ESC	ENCSR059MBO/ENCSR326ULS	ChIP for H3K27me3 and Input, E14 mESC	ENCODE	ENCODE
ChIP-seq	ChIP_H3K9me2_ESC	GSE97877	ChIP for H3K9me2, mESC	Poleshko et al., 2017	Rajan Jain
ChIP-seq	ChIP_LaminB_ESC	GSE97877	ChIP for LaminB, mESC	Poleshko et al., 2017	Rajan Jain
ChIP-seq	ChIP_Dnmt3a1_ESC	GSE100951	ChIP for Dnmt3a1, WT and Tet1 KO mESC	Gu et al., 2018	Margaret Goodell
ChIP-seq	ChIP_Dnmt3a1_ESC	GSE96527	ChIP for Dnmt3a1 and input in WT mESC	Manzo et al, EMBO J, 2017	Tuncay Baubec
ChIP-seq	ChIP_Dnmt3a2_Dnmt3b1_ESC	GSE57412	ChIP for Dnmt3a2 and Dnmt3b1 in WT mESC	Baubec et al, Nature, 2015	Dirk Schübeler
ChIP-seq	ChIP_Tet1_ESC	GSE24841	ChIP for Tet1 (N-terminal antibody) and IgG control	Williams et al, Nature, 2011	Kristian Helin
ChIP-seq	ChIP_Tet2_ESC	GSE115964	ChIP for Tet2 (N-terminal antibody) and IgG control	Rasmussen et al, Genome Research, 2019	Kristian Helin

Supplementary table S3: Datasets generated in this study

Experiment type	Experiment	Description
WGBS	WGBS_Tet2/3_DKO_NKT_expanded	WGBS in Tet2/3 DKO NKT cells expanded in immunocompetent recipients, two biological replicates
Hi-C	Hi-C_NKT_WT_thymus	Hi-C in WT thymic NKT cells
Hi-C	Hi-C_NKT_WT_spleen	Hi-C in WT splenic NKT cells
Hi-C	Hi-C_NKT_Tet2/3_DKO_NKT_expanded	Hi-C in Tet2/3 DKO NKT cells expanded in immunocompetent recipients
RNA-seq	Total_RNA_seq_NKT_WT_thymus	Ribodepleted RNA-seq in WT thymic NKT cells from young animals (3-4 week old)
RNA-seq	Total_RNA_seq_NKT_Tet2/3_DKO_young	Ribodepleted RNA-seq in Tet2/3 DKO thymic NKT cells from young animals (3-4 week old)

Supplementary table S4: Bisulfite conversion rates of methylomes analyzed in this study

Cell type	Reference	Laboratory	Genotype	Bisulfite conversion %
ESC	Yin et al, Science, 2017	Jussi Taipale	WT	99.27061
ESC	Yin et al, Science, 2017	Jussi Taipale	Tet1/2/3 TKO	98.23677
ESC	Yin et al, Science, 2017	Jussi Taipale	Dnmt1/3a/3b TKO	98.83685
ESC	Lu et al, Genome Research, 2014	Yi Zhang	WT	99.77342
ESC	Lu et al, Genome Research, 2014	Yi Zhang	Tet1/2/3 TKO	99.82283
ESC	Hon et al, Molecular Cell, 2014	Bing Ren	WT	99.60175
ESC	Hon et al, Molecular Cell, 2014	Bing Ren	Tet1 KO	99.57533
ESC	Hon et al, Molecular Cell, 2014	Bing Ren	Tet2 KO	99.70923
NKT	Tsagaratou et al, Nature Immunology, 2017	Anjana Rao	WT	99.88858
NKT	Tsagaratou et al, Nature Immunology, 2017	Anjana Rao	Tet2/3 DKO (young)	99.92096
NKT	This study	Anjana Rao	Tet2/3 DKO (expanded)	99.92077
proB	Lio et al., eLife, 2016	Anjana Rao	WT	99.27433
proB	Lio et al., eLife, 2016	Anjana Rao	Tet2/3 DKO	99.28012

Supplementary table S5: Differential expression analysis of methylation-related genes

	Tet1 vs WT, mESC, Hon et al. 2014			Tet2 vs WT, mESC, Hon et al. 2014			TetTKO vs WT, mESC, Lu et al. 2014			Tet2/3 DKO vs WT, thymic NKT (young), Tsagaratou et al. 2017			Tet2/3 DKO vs WT, splenic NKT (expanded), Tsagaratou et al. 2017			Tet2 KO vs WT, HSPC, Zhang et al. 2016		
Genes	log2FoldChange	Pvalue	Adjusted Pvalue	log2FoldChange	Pvalue	Adjusted Pvalue	log2FoldChange	Pvalue	Adjusted Pvalue	log2FoldChange	Pvalue	Adjusted Pvalue	log2FoldChange	Pvalue	Adjusted Pvalue	log2FoldChange	Pvalue	Adjusted Pvalue
Dnmt1	-0.3832762	0.3781573	0.9990316	-0.09535718	0.8465288	0.9870463	0.007848649	0.989245	0.9973187	0.01416653	0.9669503	0.994908	1.089568	0.01240315	0.09766158	0.5402262	0.005055256	0.02493475
Dnmt3a	0.2293185	0.5952186	0.9990316	0.1680226	0.7548926	0.9818612	0.1354289	0.8188068	0.9973187	0.07421201	0.8496613	0.993813	1.201698	0.0757458	0.3363919	0.1922322	0.38997	0.5784337
Dnmt3b	Not detected	Not detected	Not detected	Not detected	Not detected	Not detected	0.8917121	0.1314702	0.9973187	Not detected	Not detected	Not detected	Not detected	Not detected	Not detected	0.5088264	0.03237024	0.1020628
Tet1	KO	KO	KO	0.1808895	0.5833723	0.9609881	KO	KO	KO	Not detected	Not detected	Not detected	Not detected	Not detected	Not detected	0.1598554	0.5916572	0.7458601
Tet2	0.1941287	0.6563286	0.9990316	KO	KO	KO	KO	KO	KO	KO	KO	KO	KO	KO	KO	KO	KO	KO
Tet3	-0.07295958	0.8554356	NA	-0.05068782	0.9251563	NA	KO	KO	KO	KO	KO	KO	KO	KO	KO	0.6914659	0.0204745	0.07246841
Idh1	0.6290385	0.08422165	NA	Not detected	Not detected	Not detected	0.1277666	0.8284318	0.9973187	Not detected	Not detected	Not detected	Not detected	Not detected	Not detected	0.2519771	0.3018423	0.4913999
Idh2	0.2216031	0.610965	0.9990316	0.2409174	0.6491521	0.9640058	Not detected	Not detected	Not detected	Not detected	Not detected	Not detected	Not detected	Not detected	Not detected	0.4151734	8.29E-06	0.0001229698
Uhrf1	1.047607	0.0158462	0.4789873	1.161828	0.02476095	0.4002242	-0.02180673	0.9682552	0.9973187	0.2084523	0.6176339	NA	2.858565	0.0007859357	0.01117721	0.6577167	0.00365927	0.01924505

Supplementary table S6: Summary of DNA methylation changes in TET-deficient models

TET-deficient models

Cell type	Compartment	Comparison	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
HSC	Euchromatin	Tet2 KO%-WT%	-46.87	-2.699	-0.4847	-0.4456	1.719	36.18
HSC	Heterochromatin	Tet2 KO%-WT%	-44.97	-5.712	-2.134	-2.451	1.105	31.63
ESC (Yin, 2017)	Euchromatin	Tet TKO%-WT%	-25.46	-4.114	-1.75	-1.27	1.046	46.12
ESC (Yin, 2017)	Heterochromatin	Tet TKO%-WT%	-28.33	-7.683	-4.906	-5.135	-2.439	30.55
ESC (Hon, 2014)	Euchromatin	Tet1 KO%-WT%	-64.55	-12.36	-9.746	-9.661	-6.95	44.8
ESC (Hon, 2014)	Heterochromatin	Tet1 KO%-WT%	-58.62	-17.16	-13.9	-13.98	-10.76	31.28
ESC (Hon, 2014)	Euchromatin	Tet2 KO%-WT%	-69.55	-11.55	-7.923	-7.469	-3.639	49.84
ESC (Hon, 2014)	Heterochromatin	Tet2 KO%-WT%	-70.5	-19.5	-15.89	-15.73	-12.12	39.47
ESC (Lu, 2014)	Euchromatin	Tet TKO%-WT%	-39.01	-8	-4.595	-3.932	-0.4709	48.23
ESC (Lu, 2014)	Heterochromatin	Tet TKO%-WT%	-38.51	-12.25	-8.92	-8.929	-5.681	51.18
ESC->NPC	Euchromatin	Tet2 KO%-WT%	-45.06	-1.663	-0.01705	-0.1791	1.526	66
ESC->NPC	Heterochromatin	Tet2 KO%-WT%	-60.26	-4.71	-1.92	-2.335	0.4641	39.76
NKT (young)	Euchromatin	Tet2/3 DKO%-WT%	-38.66	0.4957	2.852	3.313	5.759	47.5
NKT (young)	Heterochromatin	Tet2/3 DKO%-WT%	-29.31	-8.386	-5.889	-5.717	-3.314	26.29
NKT (expanded)	Euchromatin	Tet2/3 DKO%-WT%	-41.28	0.8537	4.264	4.729	8.743	66.02
NKT (expanded)	Heterochromatin	Tet2/3 DKO%-WT%	-62.43	-19.64	-14.86	-14.12	-9.627	41.52
proB cells	Euchromatin	Tet2/3 DKO%-WT%	-21.15	-0.9866	0.1347	0.2238	1.336	29.64
proB cells	Heterochromatin	Tet2/3 DKO%-WT%	-16.19	-3.607	-1.804	-1.768	-0.000231	27.31

Human data (data from ref. 22)

Source	Cancer type	Patient	Highly Methylated Domains	Partially Methylated Domains
BLUEPRINT consortium	T cell Prolymphocytic Leukemia	S016KWU1	73.6628916	58.6744841
BLUEPRINT consortium	T cell Prolymphocytic Leukemia	S016MSU1	76.2613685	52.8102417
BLUEPRINT consortium	T cell Prolymphocytic Leukemia	S016KWU1	67.8747499	37.7524849

References

1. Y. Xi, W. Li, BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10**, 232 (2009).
2. X. Lin *et al.*, BSeQC: quality control of bisulfite sequencing experiments. *Bioinformatics* **29**, 3227-3229 (2013).
3. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).
4. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).
5. M. Lienhard, C. Grimm, M. Morkel, R. Herwig, L. Chavez, MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics* **30**, 284-286 (2014).
6. I. Hiratani *et al.*, Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol* **6**, e245 (2008).
7. D. Peric-Hupkes *et al.*, Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* **38**, 603-613 (2010).
8. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
9. Y. Jin, O. H. Tam, E. Paniagua, M. Hammell, TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593-3599 (2015).
10. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
11. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010).
12. K. Cibulskis *et al.*, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219 (2013).
13. F. Blokzijl, R. Janssen, R. van Boxtel, E. Cuppen, MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med* **10**, 33 (2018).
14. K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
15. M. Ko *et al.*, Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc Natl Acad Sci U S A* **108**, 14566-14571 (2011).
16. M. Ko *et al.*, TET proteins and 5-methylcytosine oxidation in hematological cancers. *Immunol Rev* **263**, 6-21 (2015).
17. P. P. Lee *et al.*, A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity* **15**, 763-774 (2001).
18. C. Plessy *et al.*, clonotypeR–high throughput analysis of T cell antigen receptor sequences. *bioRxiv*. doi: 10.1101/028696 (2015).