Supplemental information

Loss of *TET2* in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis

Kasper D. Rasmussen, Guangshuai Jia, Jens V. Johansen, Marianne T. Pedersen, Nicolas Rapin, Frederik O. Bagger, Bo T. Porse, Olivier A. Bernard, Jesper Christensen and Kristian Helin

Inventory of supplemental information:

Figure S1. Rasmussen et al

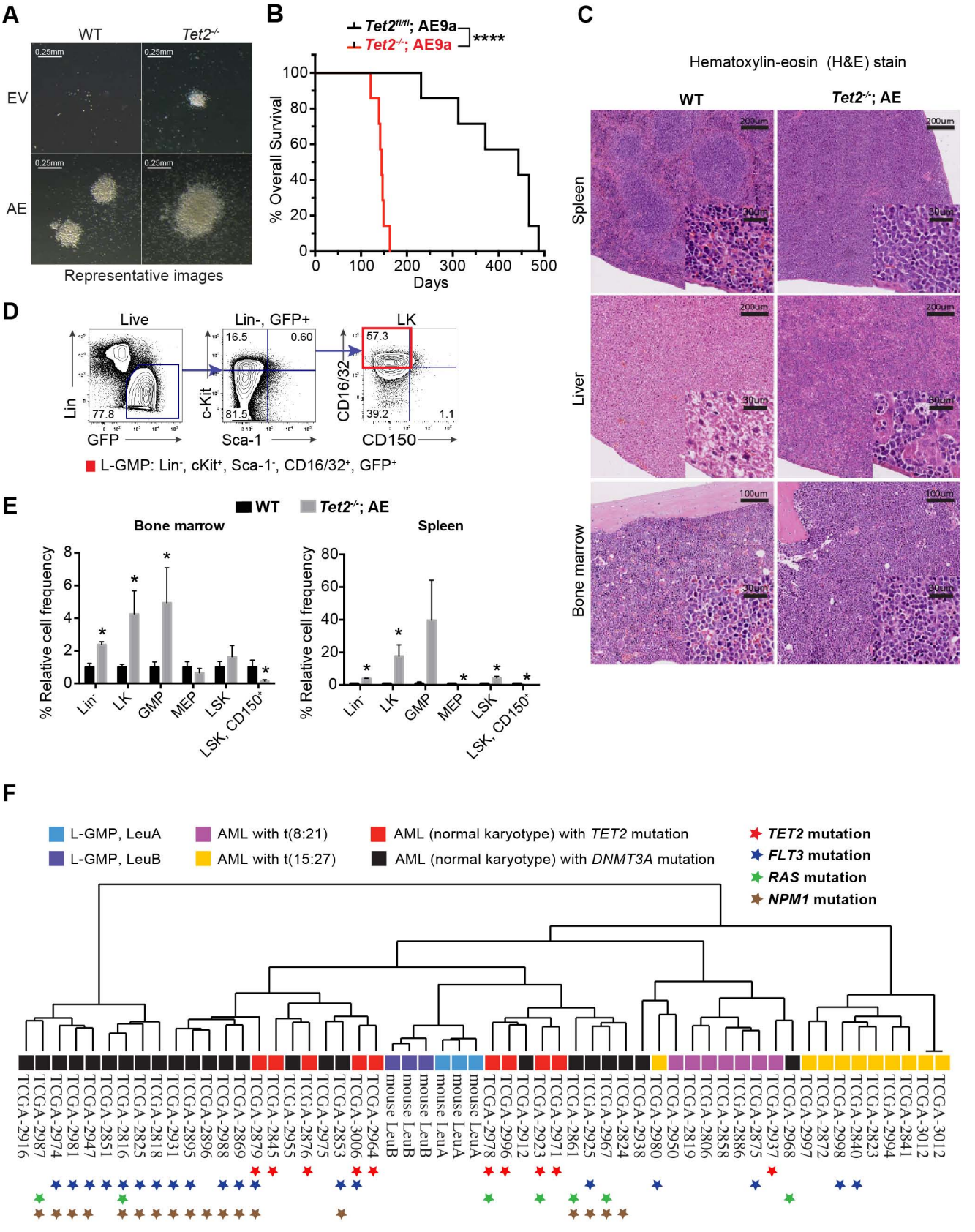


Figure S2. Rasmussen et al.

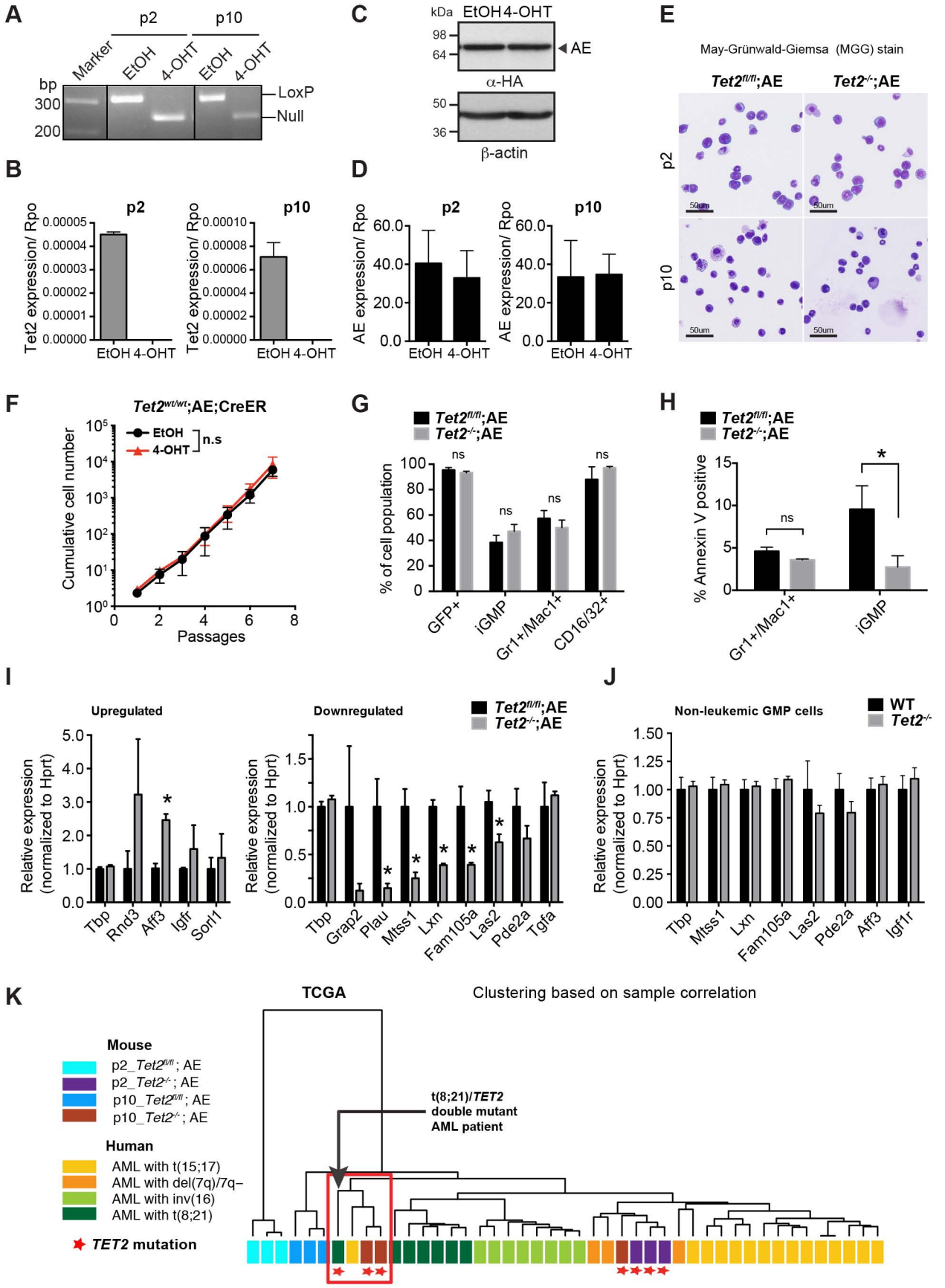


Figure S3. Rasmussen et al.

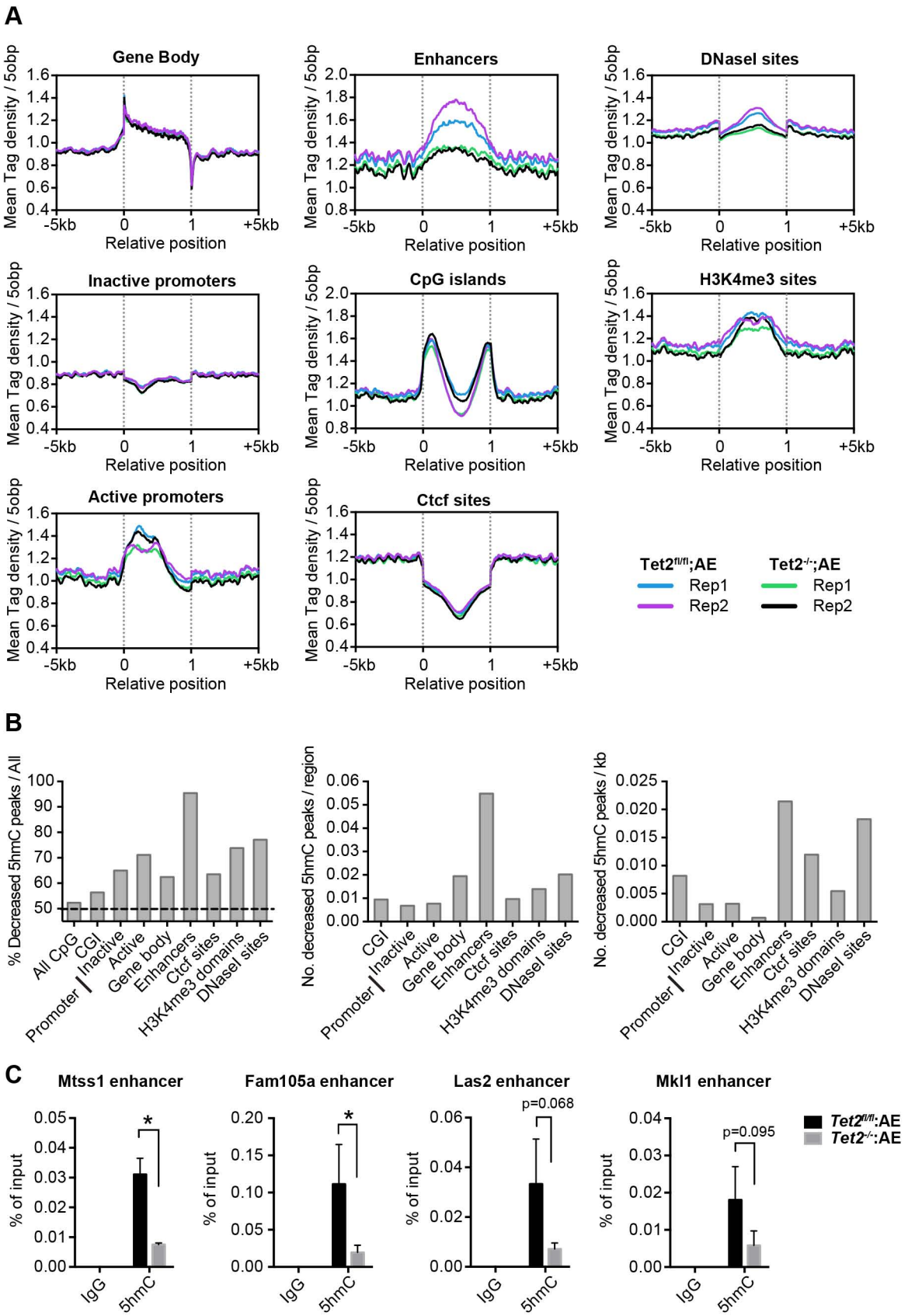
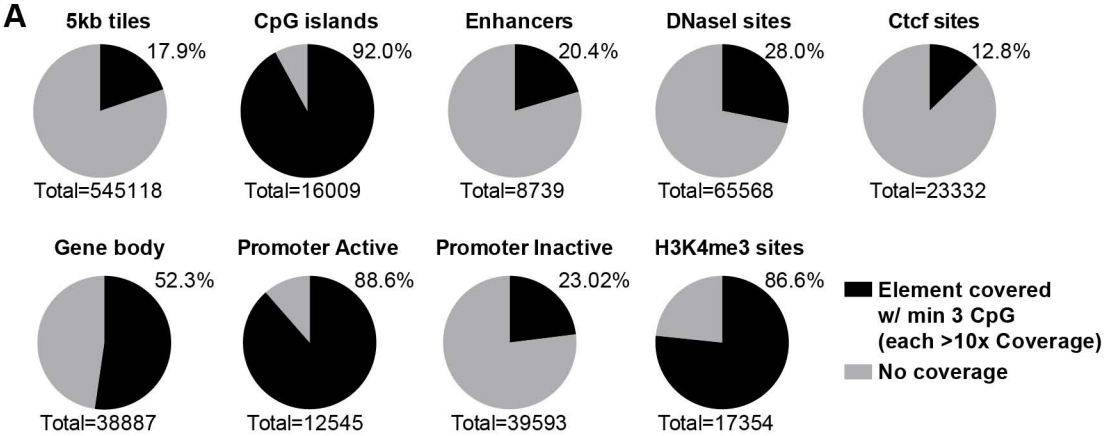
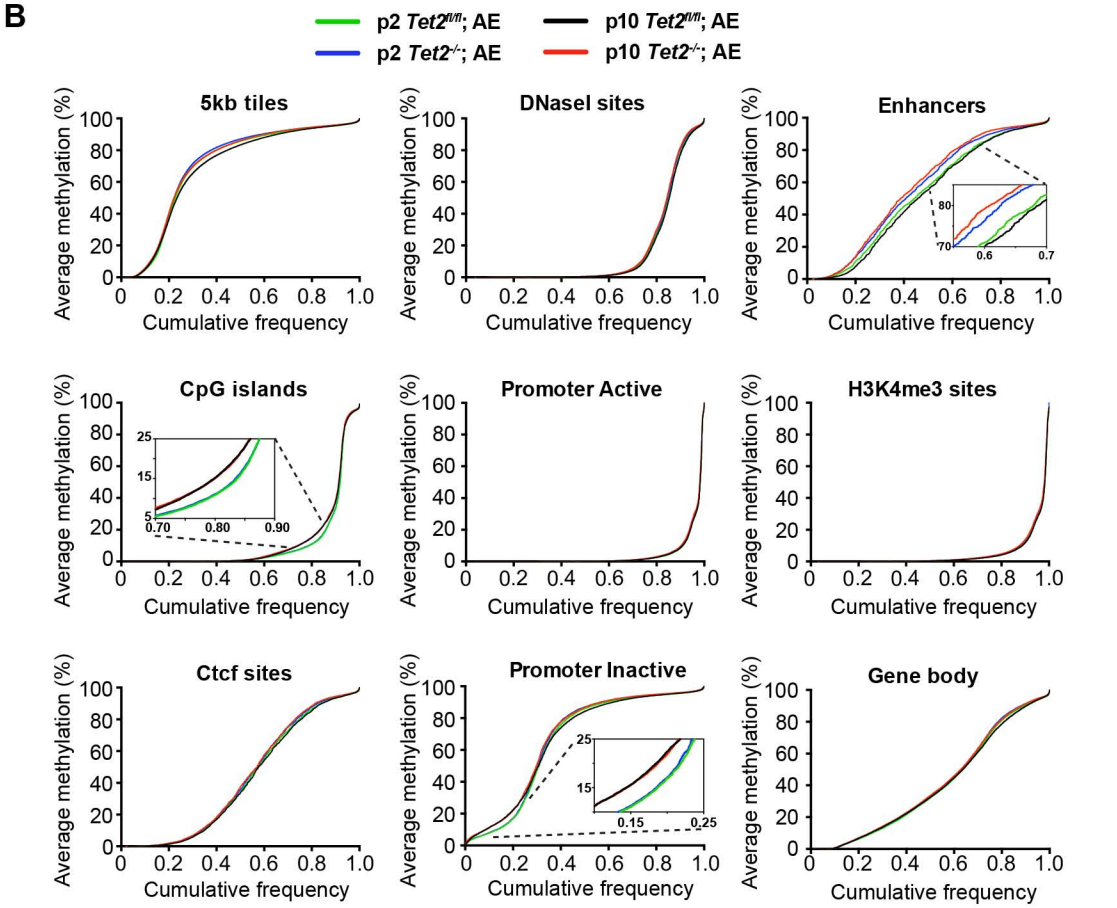


Fig S4. Rasmussen et al.





C

Effect size of DNA methylation changes (related to figure 4C and 4D)

	WT		p2		p10	
	Cohens d	effect size r	Cohens d	effect size r	Cohens d	effect size r
All CpG	-0.099	-0.050	-0.323	-0.160	-0.247	-0.122
CpG island	-1.082	-0.476	-0.060	-0.030	0.171	0.085
Promoter - Inactive	-0.872	-0.400	-0.196	-0.097	0.134	0.067
Promoter - Active	-0.540	-0.261	-0.185	-0.092	-0.145	-0.072
Gene Body	-0.255	-0.126	-0.299	-0.148	-0.194	-0.097
Enhancers	0.327	0.162	-1.011	-0.451	-1.105	-0.484
Ctcf sites	-0.036	-0.018	-0.270	-0.134	-0.339	-0.167
H3K4me3	-0.453	-0.221	-0.260	-0.129	-0.228	-0.113
DNaseI sites	-0.150	-0.075	-0.390	-0.192	-0.413	-0.202

* Cohen d between 0.5 and 1.0

** Cohens d above 1.0

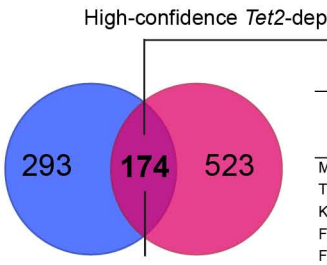
Effect size of DNA methylation change (related to Figure 4G)

	MLL-AF9 cells (p20)	
	Cohens d	effect size r
All CpG	-0.195	-0.097
CpG island	-0.261	-0.129
Promoter	-0.261	-0.129
GeneBody	-0.202	-0.101
Enhancers	-0.729	-0.342
Ctcf sites	-0.291	-0.144
DNaseI sites	-0.366	-0.180

* Cohen d between 0.5 and 1.0

Figure S5. Rasmussen et al.

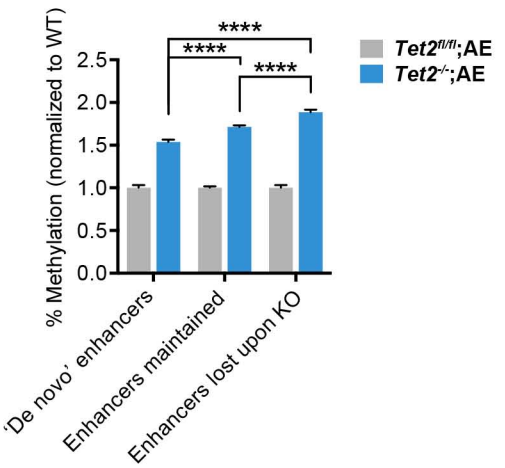
A



Nearest gene, FDR<0.05

Gene	Log2 fold change (p10)	FDR	Gene distance
Mtss1	-1.19	0.00136	0
Tgfa	-1.22	0.0118	0
Kit	-0.81	0.03	69811
Fam105a	-0.52	0.0307	11321
Fam105a	-0.52	0.0307	0
Igf1r	0.67	0.0323	0

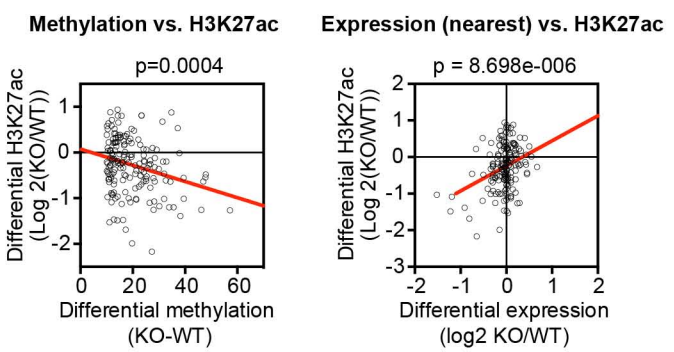
B



C

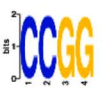
Pearson's correlation on high confidence *Tet2*-dependent enhancer-DMRs

	Differential methylation (KO-WT)	Differential H3K27ac (Log 2(KO/WT))	Differential 5hmC Log2(KO/WT)	Differential expression (log2 KO/WT)
Differential methylation (KO-WT)		-0.264	0.061	-0.059
Differential H3K27ac (Log 2(KO/WT))	-0.264		-0.079	0.330
Differential 5hmC Log2(KO/WT)	0.061	-0.079		-0.051
Differential expression (log2 KO/WT)	-0.059	0.330	-0.051	



D

Enriched TF motifs in hypermethylated enhancer-DMRs

Motif	E-value	TF
	5.9e-012	ELK4
	2.3e-004	SPI1 FEV SPIB
	2.3e-002	Runx1

E

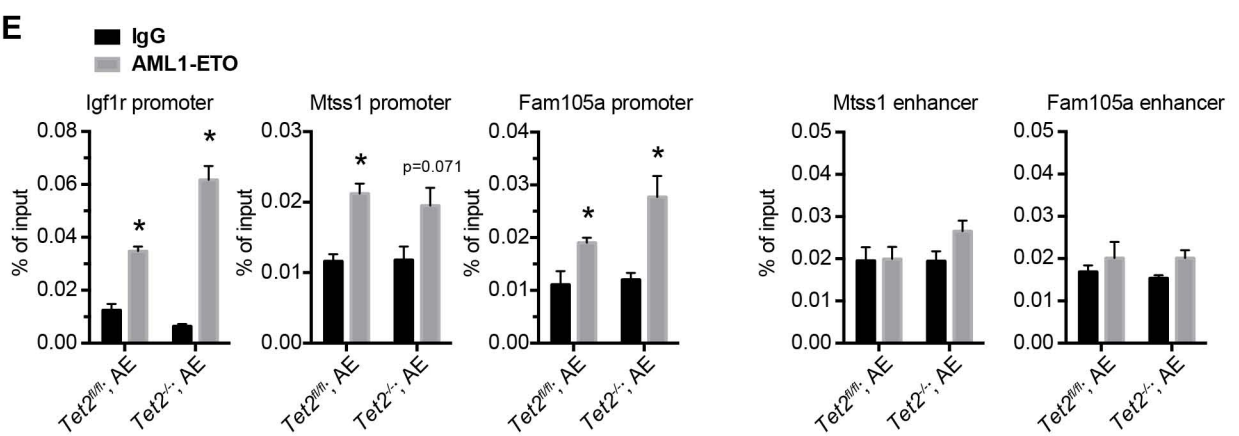
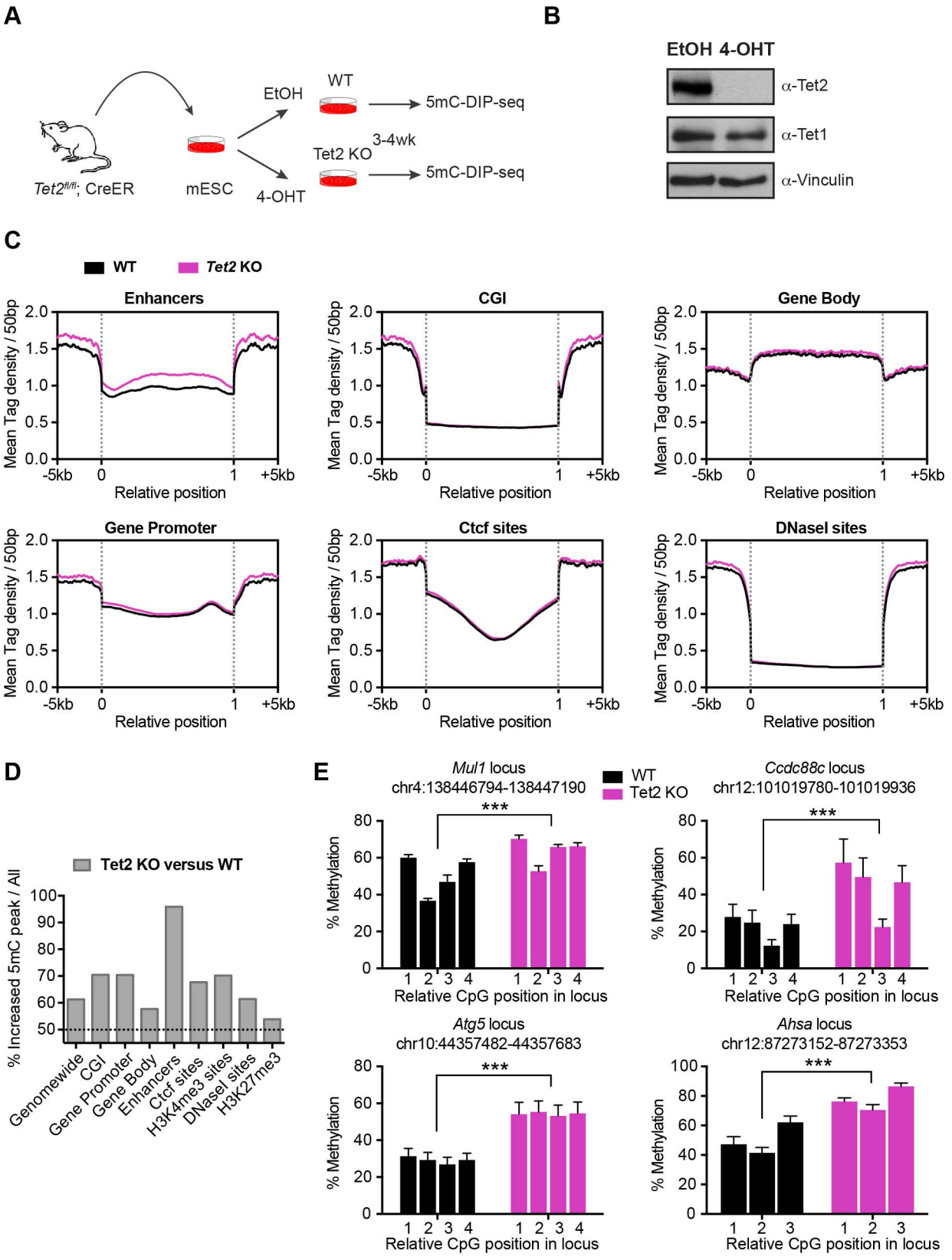


Figure S6. Rasmussen et al



Supplemental Figure Legends

Figure S1. Loss of *Tet2* and AML1-ETO expression collaborate to induce Acute Myeloid Leukemia

(A) Serial replating assay of *Tet2*$^{fl/fl}$ or *Tet2*$^{-/-}$ Kit-enriched HSPCs transduced with either empty vector (EV) or AML1-ETO (AE) expressing retrovirus. Representative images of colonies observed in serial replating assays after 5 replatings in methylcellulose-containing media. Colonies formed by *Tet2*$^{-/-}$;AE cells are significantly larger than those formed by *Tet2*$^{-/-}$:EV or *Tet2*$^{fl/fl}$;AE.

(B) Kaplan-Meier plot showing overall survival of lethally irradiated (900 Rad) recipient SJL mice transplanted with *Tet2*$^{fl/fl}$ (n=7) or *Tet2*$^{-/-}$ (n=7) Kit-enriched HSPCs transduced with AE9a expressing retrovirus. ****, p-value<0.0001 (Wilcoxon test).

(C) Hematoxylin and eosin (HE) sections of spleen (Upper panel), liver (Middle panel) and bone marrow tissues (Lower panel) obtained from WT or moribund mice transplanted with *Tet2*$^{-/-}$; AE HSPCs. The images and magnified inserts are representative of biological duplicate experiments.

(D) Representative FACS-analysis and gating strategy of major leukemic population in the bone marrow of moribund recipient mice 1 month after transplantation with *Tet2*$^{-/-}$;AE leukemic cells. GFP-positive leukemic blasts are abundantly present in the bone marrow of diseased mice and show signs of myeloid origin (Lin-cKit+Sca1-CD16/32+CD150-) (Pronk et al., 2007).

(E) Comparative enumeration of hematopoietic progenitors in bone marrow (left panel) and spleen (right panel) isolated form WT mice (n=4) or moribund mouse transplanted with *Tet2*$^{-/-}$;AE cells (n=4) as in C. Lin^{-};Lineage-negative cells. LK; Lineage-negative, Kit-positive. GMP; Granulocyte-Monocyte-Progenitor. MEP; Megakaryocyte-Erythrocyte-Progenitor. LSK; Lineage-negative, Sca1-positive, Kit-positive. LSKCD150^{+}; LSK cells, CD150-positive (Pronk et

al., 2007). The frequency of each cell population was normalized to WT. Bars represent mean, error bar indicate SD. *, p-value<0.05 (Student's t-test).

(F) Unsupervised hierarchical clustering of leukemic GMPs (LeuA and LeuB) together with human AML samples from TCGA (Cancer Genome Atlas Research Network, 2013). For clarity, TCGA patients with co-occurring mutations in multiple genes involved in DNA methylation (*TET1*, *TET2*, *DNMT1*, *DNMT3A*, *DNMT3B*, *IDH1* and *IDH2*) were excluded and only patients with *TET2* mutations, *DNMT3A* mutations, t(8;21) translocations, as well as an APL control group are shown. Patients with *TET2*, *NPM1*, *RAS* and *FLT3* mutations are indicated with asterisks. Patients with *TET2* mutations cluster consistently close to L-GMP samples from mice, whereas the cytogenetically normal *DNMT3A*-mutated group shows a more promiscuous clustering pattern.

Figure S2. Disruption of *Tet2* in pre-leukemic AE cell cultures leads to gene expression changes present in human AML with *TET2* mutations

(A) Complete recombination of *loxP*-flanked *Tet2* allele was observed at early (p2) and late (p10) passage upon treatment with 4-OHT. In addition, we did not observe any leakiness of the inducible Cre recombinase even at late time points.

(B) Undetectable levels of *Tet2* mRNA after induction of Cre activity by 4-OHT. Primers were directed against the C-terminal exon 11, which becomes excised upon Cre induction.

(C) No change in AE protein expression upon Tet2 disruption. The figure shows western blot against the HA-tagged AE oncofusion protein and β-actin loading control on cell cultures at passage 5 after treatment with EtOH control and 4-OHT.

(D) No change in AE mRNA level upon Tet2 disruption. Histogram represents qPCR analysis of cell cultures (n=3) with primers directed against the AE fusion protein.

(E) No observed change in morphology at early or late passage after *Tet2* disruption. Representative images of May-Grünwald-Giemsa (MGG)-stained cytospins of *Tet2*$^{fl/fl}$;AE and *Tet2*$^{-/-}$:AE cultures at passage 2 and passage 10. Bars, 50μm.

(F) Off-target effects of the Cre recombinase are not responsible for the acceleration of cell growth observed upon deletion of *Tet2*. AE transduced cell cultures (n=2) were derived from CreER mice with wild type *Tet2* alleles and treated with EtOH or 4-OHT. No change in proliferation was observed even after extended passaging. n.s., not significant.

(G) Disruption of *Tet2* does not change the fraction of iGMP and mature granulocytes in cultures of *in vitro* preleukemic cells even after extended cell passaging. Histograms show mean fraction (n=3) of GFP-positive, iGMP, Gr1+Mac1+, as well as CD16/32+ cells as measured by flow cytometry in *Tet2*$^{fl/fl}$:AE and *Tet2*$^{-/-}$:AE cells at passage 10. Data are representative of two independent experiments. *, $p<0.05$ (Student's t-test).

(H) Cell-type specific decrease of apoptotic rate observed in *Tet2*$^{-/-}$:AE iGMP cells. Histogram show mean fraction (n=3) of Annexin V positive cells in either Gr1+Mac1+, or iGMP cell populations in *Tet2*$^{fl/fl}$:AE and *Tet2*$^{-/-}$:AE cultures at passage 10. Data are representative of two independent experiments. *, $p<0.05$ (Student's t-test).

(I) qRT-PCR analysis of selected deregulated genes in independently derived *Tet2*$^{fl/fl}$;AE and *Tet2*$^{-/-}$;AE cultures grown for 5 passages after Tet2 disruption. Bars represent mean values (n=3), error bars indicate SEM. *, p-value<0.05 (Student's t-test).

(J) qRT-PCR analysis of selected deregulated genes (from I) in sorted non-leukemic GMP cells from 12 week-old wildtype and Tet2$^{-/-}$ mice. Bars represent mean values (n=4), error bars indicate SEM.

(K) Gene expression changes upon *Tet2* deletion and extended cell passaging are highly related to a human AML patient with a co-occuring t(8;21) translocation and *TET2* mutation. The dendrogram represents sample correlation-based clustering of preleukemic *Tet2*$^{fl/fl}$:AE and *Tet2*$^{-/-}$:AE cells grown for 2 or 10 passages after Tet2 disruption and patients with

different karyotypic aberrations from the cancer genome atlas dataset (Cancer Genome Atlas Research Network, 2013). The patient harboring the co-occuring t(8;21)/*TET2* mutation is indicated with an arrow.

Figure S3. 5hmC is specifically lost at enhancers elements

(A) Summarized 5hmC-DIP-seq read densities across various genomic elements. A distinct and reproducible decrease in read densities can be observed in enhancers (defined in this study) and DNaseI sites (as defined in MEL cells), whereas there is little or no change in other elements. Each graph shows the sample-normalized summarized read densities in the biological duplicate samples (Rep1 and Rep2) of $Tet2^{fl/fl}$:AE and $Tet2^{-/-}$:AE cultures relative to the length of the genomic element as well as 5kb of flanking regions.

(B) Quantitative peak analysis using "MEDIPS" (version 1.8.0) R-package on replicate 5hmC-DIP-seq samples of $Tet2^{fl/fl}$:AE and $Tet2^{-/-}$:AE cells shows specific loss of 5hmC at enhancers. Only peaks that were consistently changing (q-value<0.05, Log2 fold change<0.5) in the biological duplicate samples (Rep1 and Rep2) were analyzed. (Left) Ratio of down-regulated 5hmC peaks versus all significantly changing 5hmC peaks upon *Tet2* disruption, within the indicated genomic elements. On a genome-wide scale (All CpG), 52.3% of the peaks went down and 47.7% went up, indicating a nearly stochastic distribution. In contrast, when we interrogated enhancer elements we found that 95.4% of the peaks were reduced upon loss of *Tet2*. (Middle) Number of significantly down-regulated peaks upon *Tet2* disruption normalized to the number of regions of each genomic element. (Right) Number of significantly down-regulated peaks upon *Tet2* disruption normalized to the genomic space (kb) of each genomic element.

(C) 5hmC-DIP-qPCR validation of 5hmC depletion on selected enhancers using independently derived $Tet2^{fl/fl}$:AE and $Tet2^{-/-}$:AE cultures (n=3). See supplemental table 5 for primer

sequences. Bars represent mean enrichment over input, error bars indicate SD. *, p-value<0.05 (Student's t-test).

Figure S4. Enhanced reduced representation bisulfite sequencing reveals preferential DNA hypermethylation in enhancer elements

(A) Pie charts showing the eRRBS read coverage of individual genomic elements based on the 2.3 million individual CpG sites assayed. A genomic element was considered to be covered if it contained at least 3 individual CpGs, each covered by more than 10 reads in all 4 conditions. eRRBS enriches the CpG-rich regions of the genome (Akalin et al., 2012). Consistently, 92.0% of CpG islands were covered in this dataset. In addition, we could show that eRRBS was able to cover a considerable amount of enhancers (~20%) in our system. The total number of genomic elements in each category is indicated below each pie chart and the covered fraction is indicated by a percentage.

(B) Cumulative distribution plots showing the average DNA methylation of various genomic elements. Only covered elements (defined in A) were considered and the average DNA methylation of each element was calculated and plotted to obtain a comparable measure. CpG islands and Inactive promoters show a passage-dependent increase in DNA methylation (See insert; p10 higher than p2 samples). Enhancers show a progressive *Tet2*-dependent increase in DNA methylation levels (See insert; *Tet2*$^{-/-}$;AE higher than *Tet2*$^{fl/fl}$;AE samples).

(C) Table showing effect size of DNA methylation change related to Figure 4C-D and 4G. *Cohen's d* was calculated using the mean and standard deviation of the differential DNA methylation levels of individual CpGs changing significantly (q-value<0.05; abs(diff)>20%) within each genomic element in *Tet2*$^{-/-}$;AE versus *Tet2*$^{fl/fl}$;AE cells. A large numeric value (e.g. *Cohens d* > ±1.0) is indicative of a strong effect.

Figure S5. Loss of 5hmC and gain of methylation leads to decrease of H3K27 acetylation and expression of neighboring genes

(A) Venn diagram showing significant (p=0.034, Fisher's test, Yates corrected) overlap of enhancer-DMRs with increased DNA methylation (>10% increase) and decreased 5hmC (Log2 fold depletion >0.5) upon *Tet2* disruption. 174 of the 467 hypermethylated enhancer-DMRs are also depleted of 5hmC, thus representing a high-confidence list of hypermethylated Tet2-dependent enhancers. Extraction of the closest genes that are significantly deregulated (FDR<0.05) reveals predominant downregulation of gene expression (Right panel, See also Table S3). In contrast, a sizeable amount of enhancer-DMRs show loss of 5hmC without being hypermethylated. It is possible that i) there is a certain lag period between the loss of 5hmC and the DNMT-catalyzed increase in DNA methylation and ii) that the inability of RRBS to discern 5hmC from mC could underestimate the hypermethylation events due to the concomitant loss of 5hmC.

(B) Enhancers that are lost upon disruption of *Tet2* are associated with the largest increase in DNA methylation. ChIP-seq analysis of $Tet2^{fl/fl}$:AE and $Tet2^{-/-}$:AE cells at passage 10 revealed 2278 enhancers that were lost as well as 3076 'de novo' enhancers gained upon Tet2 KO. Bars represent mean change of DNA methylation of CpG sites that showed significantly different DNA methylation levels (q-value<0.05; abs(diff)>20%) within the indicated enhancer subsets. ****, p-value<0.0001 (Student´s t-test).

(C) Table indicating Pearson's correlation coefficient for pairwise comparison of differential DNA methylation, 5hmC abundance, H3K27ac abundance and expression of nearest gene for the subset of 174 high-confidence *Tet2*-dependent enhancers from A (see also supplemental table S3). XY-scatter plot and linear regression (red line) of pairwise datasets of H3K27ac abundance versus differential methylation (lower left) and gene expression of nearest gene (lower right).

(D) Analysis of enriched transcription factor binding motifs in the 467 hypermethylated enhancer-DMRs. The highest scoring 'CCGG' motif is likely an artifact related to the biased eRRBS coverage of CpG-rich regions (related to the target specificity of the MspI restriction enzyme).

(E) Binding of AE to promoters of deregulated genes is not significantly altered upon disruption of Tet2. Bar charts represent enrichment of AE compared to IgG control on selected loci as measured by ChIP-qPCR at passage 5 after Tet2 disruption. AE binding cannot be detected in associated Tet2-dependent enhancers. Bars represent mean enrichment over input, error bars indicate SEM. *, p-value<0.05 (Student's t-test).

Figure S6. Disruption of *Tet2* in embryonic stem cells is associated with enhancer hypermethylation.

(A) Overview of experimental setup. Independent embryonic stem cell lines were derived from blastocysts isolated from mice with a *Tet2*$^{fl/fl}$:CreER genotype. After 4-OHT mediated disruption of *Tet2*, the cells were grown for 3-4 weeks and genomic DNA were isolated for DNA methylation analysis by 5mC-DIP-seq and quantitative bisulfite pyrosequencing. The growth rate of ES cells was not affected by *Tet2* disruption (data not shown).

(B) Western blot of EtOH and 4-OHT-treated ES cells showing depletion of full-lenght Tet2 protein as well as Tet1 and Vinculin loading controls.

(C) Summarized 5mC-DIP-seq read densities in ES cells with wildtype or *Tet2* KO across various genomic elements. A distinct increase in read densities can be observed in enhancers, whereas there is little or no change in other elements. Read densities relative to the length of the genomic element as well as flanking regions extending 5kb upstream and 5kb downstream are shown.

(D) Quantitative peak analysis using "MEDIPS" (version 1.8.0) R-package on biological duplicate 5mC-DIP-seq samples of wildtype and *Tet2* KO ES cells. The bar chart shows the

ratio of up-regulated peaks versus all significantly changing peaks within the indicated genomic elements. Only peaks that were consistently changing (q-value<0.05, Log2 fold change<0.5) in the biological duplicate cultures were analyzed.

(E) Quantitative bisulfite pyrosequencing validation of DNA methylation changes at four enhancers (*Mul1*, *Ccdc88c*, *Atg5*, and *Ahsa* locus) associated with gain of DNA methylation in *Tet2* KO ES cells (See supplemental table S4 for primer sequences). Bar graphs show methylation levels of each individual CpG within the locus. Bars represent mean methylation (n=3), error bars indicate SEM. ***, p-value<0.0001 (two-way ANOVA).

.

Supplemental Experimental Procedures

Mice and transplantation experiments. The conditional *Tet2* mouse line, generated as described previously (Quivoron et al., 2011), was crossed to Mx1-Cre (Kühn et al., 1995) and Rosa26-Cre-ERT2 (Ventura et al., 2007) mouse lines. Hereafter referred to as *Tet2*$^{fl/fl}$; Mx1-Cre or *Tet2*$^{fl/fl}$; CreER, respectively. Embryonic stem cells with a *Tet2*$^{fl/fl}$:CreER genotype were derived, transduced with a vector containing puromycin resistance and cultured in serum-free '2i' medium essentially as previously described (Pedersen et al., 2014) (Williams et al., 2011). To induce the acute inactivation of *Tet2 in vivo*, *Tet2*$^{fl/fl}$ (control) or *Tet2*$^{fl/fl}$:Mx1-Cre (experimental) mice were injected intraperitoneally with 250ug polyinosinic-polycytidylic acid (Amersham) dissolved in 1xPBS at days 0, 2 and 4. The mice were sacrificed 2 weeks after last injection and bone marrow extracted, enriched for Kit expression (CD117 microbeads, Miltenyi Biotech) and maintained in culture in 1× StemPro®-34 SFM (Invitrogen), 1x L-Glutamine with SCF (CHO producer cell line), 10ng/ml IL-3 (Peprotech), 10ng/ml IL-6 (Peprotech) and 0.1mM 2-mercapto-ethanol. After 2 days of pre-stimulation, the Kit-enriched hematopoietic stem and progenitor cells (HSPCs) were transduced for 2 consecutive days on RetroNectin®-coated (TaKaRa Bio) culture plates pre-incubated for 1h with viral supernatants. For primary transplantation experiments, lethally irradiated (900 Rad) recipient mice (Ly-5.1) were transplanted with 3×10^6 HSPCs (Ly-5.2) as well as 2×10^5 whole bone marrow cells (Ly5.1) for radioprotection. For the secondary transplantation experiment, sublethally irradiated (650Rad) recipient mice (Ly-5.1) were transplanted with 1×10^6 splenocytes harvested from primary moribund leukemic mice. After irradiation, all mice were maintained on medicated water (0.1g/L Ciprofloxacin) for two-three weeks.

Retroviral constructs and virus production. The MigR1-AE-IRES-GFP (Yan et al., 2006), MigR1-AE9a-IRES-GFP (Yan et al., 2006), and MSCV-MLL-AF9-IRES-Neo (Somervaille and

Cleary, 2006) retroviral constructs were co-transfected with pCL-ECO packaging plasmid into Phoenix-ECO cells using the $CaPO_4$ transfection method. Viral supernatants were harvested and filtered at day 2 and 3 after transfection and used immediately for infection of primary hematopoietic cells.

Flow cytometry. Single cell suspensions from bone marrow and peripheral blood of transplanted mice were treated with ACK lysis buffer (150mM NH_4Cl, 1mM KCO_3, 0.1mM Na_2EDTA, pH 7.3) to lyse red blood cells, washed, and resuspended in 1× PBS plus 2% FCS. 1 × 10^6 spleen or bone marrow cells, or *Tet2*$^{fl/fl}$;AE;CreER cells from *in vitro* cultures, were stained with antibody cocktails for surface markers of various hematopoietic cells. Hematopoietic stem and progenitor cells: Biotinylated lineage cocktail (CD3e, CD11b, Gr1, B220, Ter119, eBioscience), Streptavidin-PECy5 (eBioscience), Sca1-PE (D7, eBioscience), CD117-APC-eFlour®780 (2B8, eBioscience), CD16/32-PECy7 (93, eBioscience), CD150-APC (TC15-12F12.2, Biolegend). Peripheral blood cells: CD45.1-APC (A20, Biolegend), CD45.2-PE (104, Biolegend), Ter119-PECy5 (Ter119, eBioscience). *Tet2*$^{fl/fl}$;AE;CreER *in vitro* cultured cells: Gr1-APC-eFlour®700 (RB6-8C5, eBioscience), CD11b-PE (M1/70, eBioscience), CD16/32-PECy7 (93, eBioscience). Apoptosis assay: Gr1-PE (RB6-8C5, eBioscience), CD11b-PE (M1/70, eBioscience), CD16/32-PECy7 (93, eBioscience) and Annexin V-APC (eBioscience). 7AAD were added to exclude dead cells from analysis. Cells were analysed and sorted on FACSAria I (BD) FACSAria III (BD) and LSRII (BD) in the Flow Cytometry Facility at BRIC, University of Copenhagen and analyzed using the FlowJo© software (Tree Star, Inc.).

Histological analysis of tissues and peripheral blood. To assess morphology, peripheral blood or single-cell suspensions of *in vitro* cultured cells was smeared or spun onto glass slides, air-dried and stained with May-Grünwald-Giemsa (Merck) according to manufacturers instructions. Femur, spleen and liver were dissected from mice and fixed 48h

in Lillie's solution (10% formalin). Femur samples were decalcified in 10% EDTA, pH 7.4, and all tissues were dehydrated prior to embedding in paraffin for sectioning. All sections were stained with Hematoxylin and Eosin (HE) according to manufacturers instructions (Sakura Finetek). Images of tissues and blood samples were acquired using the NanoZoomer Digital Pathology System (Hamamatsu).

Colony formation assay. Transduced HSPCs were plated in methylcellulose medium (M3534, StemCell Technologies) according to manufacturer's protocol. Cells were seeded with equal density (10000 cell/plate) except *Tet2*$^{-/-}$;AE cells which after the first two round of replating was seeded with a density of 2500 cells/plate to avoid overplating. Colonies were counted and replated every five to seven days.

Dot blot for measuring global 5hmC levels. Genomic DNA was isolated with DNeasy Blood & Tissue Kit (Qiagen). To denature the DNA, the samples were incubated with 0.4 M NaOH, 10 mM EDTA at 95°C for 10 min, and a final concentration of 1M of cold ammonium acetate (pH 7.0) was added for neutralization. 2-fold serial dilutions of denatured DNA were spotted on a Hybond-C membrane (Fisher/GE-Healthcare) in Bio-Dot apparatus (Bio-Rad) according to manufacturer's protocol. The membrane was washed with 2× SSC buffer [300 mM NaCl, 300mM Trisodium citrate (pH 7.0)], and baked at 80°C for 2 hours, then blocked with 5% skim milk in 1×PBS with 0.1% Triton-X 100 overnight at 4°C, and incubated with anti-5hmC antibody (Active Motif #39791, 1:1,000) for 2 hours at room temperature. After incubating with HRP-conjugated secondary antibody, the membrane was washed and visualized by enhanced chemiluminescence (ECL, Amersham Biosciences). The same membrane was hybridized with methylene blue (0.02% in 0.3M Sodium acetate solution, pH 5.3) to ensure the equal spotting amount of total DNA.

Gene expression analysis.

FACS-sorted *in vivo* GMP cells (Lin-cKit+Sca1-CD16/32+CD150-) were isolated from bone marrow of recipient mice one month after transplantation of WT bone marrow or splenic cells from two independent moribond *Tet2*$^{-/-}$:AE leukemic mice (LeuA and LeuB, see Fig 1B). Total RNA was isolated using the Allprep Micro Kit (Qiagen) and biological triplicate samples were labeled and hybridized to Agilent SurePrint G3 Mouse GE 8x60K arrays according to manufacturers instructions. Live, iGMP cells (GFP+, Gr1^{-}, Mac1^{-}, CD16/32^{+}) were sorted from *Tet2*$^{fl/fl}$;AE;CreER *in vitro* cultures treated with EtOH or 4-OHT and grown for 2 or 10 passages, respectively. Total RNA was isolated using the RNeasy Micro Kit (Qiagen), and biological triplicate samples were labeled and hybridized to GeneChip Mouse Gene ST 2.0 Arrays (Affymetrix), by the RH Microarray center at Rigshospitalet, Denmark according to manufacturer's instructions. Prior to array hybridisation on either the Agilent or Affymetrix platforms, the integrity of the total RNA was tested on the Agilent RNA 6000 Nano chip (Agilent).

Agilent microarrays were normalized using the RMA function from the Bioconductor affy package (Gautier et al., 2004). The normalized arrays were then background corrected using the normal+exponential convolution model with saddle-point approximation to maximum likelihood (Ritchie et al., 2007) and normalized between arrays by Cyclic Loess using the limma package implementation (Smyth and Speed, 2003). The Affymetrix arrays were annotated based on the annotation files folder "MoGene-2_0-st_rev1.zip" downloaded from Affymetrix, background corrected, normalized and probe set summarized (RMA) using Affymetrix Power Tools with default settings. Furthermore, APT calculated a p-value for Detection Above Background (DABG) for each probe signal. The RMA results from the APT processing were filtered to remove probes not detected above background in all samples in at least one group (DABG "present" p-value <= 0.05). Probe sets annotated as controls or cross-hybridizing were also removed. Several statistical tests of group comparisons were

made using the 'limma' R-package (version 3.18.0), which calculates moderated t-statistics, and the resulting p-values were corrected for multiple testing. Filtering, statistics and plotting was done in R (version 2.15).

For qRT-PCR, reverse transcription was performed with TaqMan® Reverse Transcription Reagents kit with random primers (Applied Biosystems). Real-time PCR were carried out with the LightCycler® 480 SYBR Green I Master (Roche) on LightCycler480 Real-Time PCR System (Roche). Expression values were normalized to the Hprt housekeeping gene using the ΔΔ-ct method. See supplemental table S4 for primer details.

Quantitative DNA methylation analysis by bisulfite pyrosequencing. PCR reactions were performed on bisulfite treated genomic DNA using the PyroMark PCR kit (Qiagen) with primer sets containing one biotinylated primer (see Table S4 for primer details). The resulting amplicon was immobilized on Streptavidin-coated Sepharose Beads (GE healthcare) and processed on the PyroMark vacuum prep workstation (Qiagen) to yield ssDNA for sequencing. Finally, pyrosequencing was performed on the PyroMark Q24 instrument (Qiagen) and CpG methylation was analysed using the PyroMark Q24 Software 2.0 (Qiagen). In each assay, whole-genome amplified genomic DNA obtained using the Illustra GenomiPhi v2 DNA amplification kit (GE healthcare) and M.SssI-treated genomic DNA (Milipore) were used as an unmethylated and fully methylated control, respectively.

5hmC/5mC DNA immunoprecipitation and sequencing. 5hmC/5mC DNA Immunoprecipitation (5hmC/5mC-DIP) of genomic DNA was performed as previously described (Williams et al., 2011). For 5hmC-DIP, genomic DNA was isolated from iGMP cells sorted from *Tet2*$^{fl/fl}$;AE and *Tet2*$^{-/-}$;AE cultures grown for 5 passages after Tet2 deletion. This time point was chosen as maximum global 5hmC depletion, as measured by dot blot, was already achieved (data not shown). For 5mC-DIP, genomic DNA was isolated from wildtype

or Tet2 KO ES cells grown for 3-4 weeks after Tet2 deletion. The genomic DNA was sonicated using the Bioruptor (Diagenode) to obtain fragments of approximately 200-500bp. Next, adaptors were ligated using the NEBNext® DNA sample Prep Master Mix according to manufacturer's instructions (New England Biolabs) and resuspended in 1× IP buffer (10mM NaPhosphate pH 7.0, 0.14M NaCl, 0.5% Triton X-100). 1ug of adaptor ligated DNA was briefly denatured and incubated 4h using an IgG control antibody or a polyclonal rabbit antibody directed against 5hmC (Williams et al., 2011) or a monoclonal 5mC antibody (Eurogentec BI-MECY-0500). Anti-Rabbit/mouse dynabeads (Invitrogen) were added and incubated for additional 2h before immunoprecipitation, washing and elution by proteinase K digestion. The eluted DNA was PCR-amplified (14x PCR cycles) using indexed multiplex primers for illumina sequencing (New England Biolabs), pooled and sequenced on HiSeq 2000 using 50bp single-end sequencing at Beijing Genomics Institute (BGI, Shenzhen, China).

Chromatin immunoprecipitation and sequencing. Chromatin immunoprecipitation and sequencing (ChIP-seq) was performed essentially as previously described (Pasini et al., 2010) using anti-H3K27ac (ab4729, Abcam), anti-H3K4me1 (39635, Active Motif), and anti-H3K4me3 (9751S, Cell signalling) as well as IgG control antibody. ChIP-qPCR on AE were performed using an AML1-ETO specific antibody (C15310197, Diagenode). In brief, duplicate cultures of *Tet2*$^{fl/fl}$;AE or *Tet2*$^{-/-}$:AE *in vitro* cells grown for 10 passages after *Tet2* disruption were fixed for 10 minutes in 1% formaldehyde dissolved in PBS and the reaction was quenched by addition of glycine to a final concentration of 0.125M. The crosslinked cells were harvested in SDS lysis buffer (100 mM NaCl, 50mM Tris-HCl (pH8.1), 5mM EDTA (pH 8.0), 0.2% NaN3, 0.5% SDS) and nuclei were pelleted and resuspended in IP buffer (100mM Tris-HCl (pH 8.6), 100mM NaCl, 5mM EDTA (pH 8.0), 0.2% NaN_3, 5% Triton X-100, 0.3% SDS). The obtained DNA was then sheared to an average size of 150-300bp (Bioruptor, Diagenode) and 40ug of solubilized chromatin was immunoprecipitated with Protein A Sepharose beads

(GE healthcare) after overnight incubation with the respective antibodies. Finally, adaptor-ligated libraries were generated from 10ng of precipitated DNA using the NEBNext® DNA library Sample Prep Master kit (New England Biolabs), PCR-amplified (15x PCR cycles) using indexed multiplex primers for illumina sequencing (New England Biolabs), and sequenced on a HiSeq2000 using 50bp single-end sequencing at the National High-Throughput Sequencing Centre (University of Copenhagen).

Enhanced Reduced Representation Bisulfite sequencing. Genomic DNA was isolated from sorted *Tet2*$^{fl/fl}$;AE and *Tet2*$^{-/-}$;AE iGMP cells grown for 2 or 10 passages, respectively. Enhanced Reduced Representation Bisulfite Sequencing libraries were generated as previously described (Akalin et al., 2012) with the exception that end-repair and adaptor ligation were performed using the NEBNext® DNA library Sample Prep Master kit (New England Biolabs). In brief, ~1ug of genomic DNA was digested with MspI enzyme overnight followed by phenol/chloroform extraction. After library generation using pre-annealed methylcytosine-containing illumina adaptors (TruSeq® DNA sample prep kit, Illumina) and gel-based size-selection, the ligated fragments were treated with bisulfite (EZ DNA methylation kit, Zymo) to convert unmethylated cytosines. Finally, the converted fragments were PCR-amplified (15x PCR cycles) and sequenced on an Illumina Hiseq2000 using 100bp single-end sequencing in the National High-Throughput Sequencing Centre (University of Copenhagen).

Computational analysis

Preprocessing, trimming and mapping of sequencing reads. Raw sequencing reads obtained from ChIP-seq and 5hmC-DIP-seq experiments were split according to the sample index, and 3'-adaptor sequences as well as low quality nucleotides were trimmed prior to mapping to the GRCm38 (mm10) mouse genome using bowtie (Langmead et al., 2009), with parameters

"-m1 --strata --best". Potential PCR duplicates were removed by only keeping a single read if multiple reads mapped to the exact same genomic sequence. Peak calling of ChIP-seq datasets was performed using MACS 1.4 (Zhang et al., 2008) with the corresponding IgG sample as control and a p-value cut-off of 1E-12. To analyze sequencing data obtained from eRRBS experiments, raw reads were preprocessed using TrimGalore 0.3.1 (--rrbs option) to remove 3'-adaptor sequences. In addition, mapping and methylation calling were performed using Bismark 0.10.1 (Krueger and Andrews, 2011) and bowtie2 using default settings. Finally, individual methylation calls in CpG-context were imported into SeqMonk 0.25.0 for further processing and analysis. All analyses of ChIP-seq, 5hmC-DIP and eRRBS sequencing data were performed on the public server of the Galaxy Project as well as a local installation of the Galaxy environment.

Annotation of genomic elements in mouse. A list of all mouse genes and TSS (coding and non-coding) as well as DNaseI hypersensitivity sites and Ctcf sites in MEL cells (ENCODE) were downloaded from the Ensembl database (Genes 75, Regulation 75, GRCm38.p2). A list of annotated mouse CpG islands was downloaded from UCSC database (GRCm38/mm10). Gene bodies were defined as the entire transcribed region from start to the end of a gene. To avoid redundancy only the longest transcript variant of each gene was used. Gene promoters were defined as a single region extending 1500bp upstream and 500bp downstream of all transcription start sites. Overlapping regions on the same strand, likely due to closely spaced transcription start sites were merged into a single promoter region. A gene promoter was annotated as active if having an overlapping H3K4me3 ChIP-seq peak in *Tet2*$^{fl/fl}$:AE passage 10 cells, whereas those without were annotated as inactive.

Active distal enhancers in iGMP cells harvested at passage 10 were defined as promoter-distal (Non-overlapping with a region extending ±2.5kb from all TSS), overlapping ChIP-seq enriched regions of H3K27ac and H3K4me1 histone modifications with low mean

H3K4me3 enrichment (>100 reads/kb) in both replicate samples. The resulting regions separated by less than 1kb were merged to avoid redundant detection. This yielded a final high-confidence map of enhancers comprising a total of 8739 in *Tet2*$^{fl/fl}$:AE cells and 9595 in *Tet2*$^{-/-}$:AE cells. To annotate enhancers in ES cells, BAM files of replicate ChIP-seq experiments performed on E14 ES cells using antibodies directed against H3K4me1, H3K4me3 and H3K27ac were downloaded from the ENCODE project at UCSC website (www.genome.ucsc.edu) and enhancers were defined as described above. Regions marked as DNaseI hypersensitivity sites, Ctcf sites, and H3K27me3 sites in ES cells were downloaded from the Ensembl database (Regulation 75, GRCm38.p2).

Finally, the defined regions of CpG islands, DNaseI sites, Ctcf sites, H3K4me3 sites, H3K27me3 sites, and enhancers were extended 250bp in each direction (Bedtools slop) to include flanking regions.

Annotation of genomic elements in human. A list of all human genes (coding and non-coding) was downloaded from the Ensembl database (Genes 75, GRCh37.p13). A list of annotated human CpG islands was downloaded from UCSC database (GRCh37/hg19). In addition, a comprehensive list of 389187 human transcription start sites (defined in (FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al., 2014)) as well as 43011 active enhancers (defined in (Andersson et al., 2014))were downloaded from the FANTOM consortium database. Gene bodies were defined as the entire transcribed region from start to the end of a gene. To avoid redundancy only the longest transcript variant of each gene was used. Gene promoters were defined as a single region extending 1500bp upstream and 500bp downstream of a transcription start site.

Comparative analysis of DNA methylation, 5hmC read abundance, H3K27ac read abundance and gene expression. To analyze the effect of DNA methylation, differentially

methylated regions (DMRs) were defined as regions within the 8739 enhancers in *Tet2*$^{fl/fl}$:AE cells covered by eRRBS with more than 3 CpG within 1kb (each CpG >10x coverage). This yielded a list of 2107 enhancer-DMRs with known methylation status in both *Tet2*$^{fl/fl}$:AE and *Tet2*$^{-/-}$;AE cells (See complete list in supplemental table S3). Quantification of the number of ChIP-seq and 5hmC-DIP-seq reads overlapping these regions was performed using seqMINER 1.3.3e (Ye et al., 2011). To obtain information from flanking regions the enhancer-DMRs were extended by 250bp in each direction. Read densities were extracted from read normalized BAM files ('select random lines' tool on Galaxy server) using the default settings of the 'Enrichment Based Method'. The density tables were exported, normalized to enhancer-DMR length and the mean log2 fold differential enrichment (KO/WT) for the replicate samples were found. For each enhancer-DMR, the closest gene were annotated using 'bedtools closest' from the BEDTOOLS v2.16.2 software (Quinlan and Hall, 2010). Alternatively, for the analysis presented in Figure 5D, a list of all genes within 100kb of each enhancer-DMR was computed and the mean z-score was assigned.

Analysis of enriched transcription factor binding motifs. Motif analysis was performed with command-line version of MEME-Chip {Machanick:2011ge} (version 4.9.1 using parameters "-meme-mod zoops -meme-minw 6 -meme-maxw 30 -dreme-e 0.05 -centrimo-score 5 -centrimo-ethresh 10") against the JASPAR core 2009 vertebrate database {PortalesCasamar:2010db}.

5hmC/5mC-DIP-seq analysis. Calculations of differential 5hmC/5mC status were done in R using the "MEDIPS" package (version 1.8.0) and the "BSgenome.Mmusculus.UCSC.mm10" (1.3.19) annotation package. The MEDIPS preprocessing of a sample included reading alignment file, counting reads in bins (bin size=50), determining chromosomal positions of all CpGs, determining coupling vector (fragment length=250), calibration, and count

normalization. The calculations of differential methylation between two samples were done in a sliding window (window=500bp, step=250bp) either genome wide or on a selected list of regions e.g. enhancers. To reduce the risk of inflated and misleading log2 fold changes due to small RPM (Reads Per Million) values, a small pseudocount (equal to the 1% quantile of all non-zero values) was added to all sample signal values and the log2 fold change recalculated. The output was finally filtered according to thresholds 1) RPM for either sample >= 80% quantile of all non-zero values in input sample, 2) FDR<=0.05 and 3) abs(log2 Fold Change) >= 0.5. Any overlapping significant regions were merged.

Heat map and summarized read densities across genomic elements were generated in seqMINER (Ye et al., 2011) using files of uniquely mapped reads normalized to constant reads numbers ('Select random lines' tool on Galaxy server) between *Tet2*$^{fl/fl}$;AE and *Tet2*$^{-/-}$;AE iGMP samples or wildtype and *Tet2* KO ES cells. The summarized read densities were exported, normalized to length of element and the mean log2 fold differential depletion (KO/WT) for the replicate samples were found.

Merging of microarray platforms. Microarrays was annotated using biomaRt interface for biomart (www.biomart.org) with ensemble ID as common identifiers. Human homologues based on sequence homology were acquired from same database. The probe-sets annotated to human genes where then used to merge the different datasets of human GMPs and AML samples (Cancer Genome Atlas Research Network, 2013; Haferlach et al., 2010; Rapin et al., 2014). For each gene alias represented in both platform, the probe-set with the highest intensity overall in each dataset was selected and its values reported. The overlap between the mouse Agilent and the Affymetrix HG U133 plus 2 was 15482 gene aliases, while the overlap between the Affymetrix MoGene ST 2.0 and Affymetrix HG U133 plus 2 was 15681 gene aliases. The merged datasets were then batch corrected using ComBat (Johnson et al., 2007), using the microarray platform as the parameter for batch effect. p-values throughout

this work are corrected using Bonferroni correction when needed. Unsupervised hierarchical clustering (Euclidian distance metric, Ward linkage) was generated with a combination of orange (Curk et al., 2005) and R, using the top 10% varying genes (variance filter) if not stated otherwise.

Analysis of methylation state in *TET2*-deficient AML patients. TCGA (The Cancer Genome Atlas) Illumina Infinium 450k methylation array level 3 data (pre-calculated and normalized beta values), patient mutational analysis data, and probe information file was downloaded from the TCGA homepage(Cancer Genome Atlas Research Network, 2013) (https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp). The beta values from all 194 patients were collected in a large matrix, and probes with many missing values across patients were removed with 396,019 probes remaining. 3 patients ("TCGA-2815","TCGA-2856","TCGA-2944") were excluded from the data set because of missing mutation information. A patient's missing probe beta values were imputed, by taking the median (non-missing) probe value of the 5 most similar patients (k-nearest neighbor method). Some patients had mutations in several genes involved in DNA methylation (*DNMT3A, DNMT3B, DNMT1, TET1, TET2, IDH1, IDH2*) and were assigned a "mix" mutation group. In addition, one patient harbored a compound t(8;21)/*TET2* mutation and was excluded from analysis. Finally, the remaining patients were stratified into an i) AML patient control group (103 patients) wild type for genes involved in DNA methylation and containing no t(8;21) translocation, ii) a *TET2*-mutated group (9 patients), iii) a t(8;21) translocation group (6 patients), and iv) a *DNMT3A*-mutated group (29 patients). To find differences between the respective groups and the AML patient control group, the data was filtered according to probe variance (>=75% variance quantile) and bimodality (Wang et al., 2009) (index>=0.4) and a pairwise Wilcoxon two-sample test performed. Using the 450k arrays, a cutoff of abs(Beta-value)>0.2 results in 99% confidence in detecting differential methylation (Asmar et al., 2013; Bibikova et al., 2011). Thus,

selecting probes with p-value<0.05 (Wilcoxon) and abs(Beta-value)>0.2, yielded final lists of differentially methylated CpGs that show robust and reproducible change between the AML control group, *TET2*-mutated patients, *DNMT3A*-mutated patients, and t(8:21) translocation patients. All variance, bimodality and statistics calculation were performed in R (version 2.15) using packages "bigmemory" (4.4.6), "DMwR" (0.4.1), "oompaBase" (3.0.1), "ClassDiscovery" (3.0.0). "oompaBase" and "ClassDiscovery" packages can be found here: http://bioinformatics.mdanderson.org/main/OOMPA:Overview. To analyze the distribution (enrichment/depletion) of hyper- and hypomethylated probes, each of the probes on the infinium array was assigned to a category (Gene body, Promoter, CpG islands and Enhancers) based on overlap with human annotation described above. Overlap of lists was performed on the public server of the Galaxy Project. Probes that did not fall into any of the above categories were defined as intergenic.

General statistical analysis. Unless otherwise stated, general statistical analysis were performed using GraphPad Prism version 6 for Mac OSX (Graphpad Software, La Jolla California, USA).

Supplemental references

Akalin, A., Garrett-Bakelman, F.E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., Milne, T.A., Huang, Y., Biswas, D., Hess, J.L., et al. (2012). Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet. *8*, e1002781.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Asmar, F., Punj, V., Christensen, J., Pedersen, M.T., Pedersen, A., Nielsen, A.B., Hother, C., Ralfkiaer, U., Brown, P., Ralfkiaer, E., et al. (2013). Genome-wide profiling identifies a DNA methylation signature that associates with TET2 mutations in diffuse large B-cell lymphoma. Haematologica *98*, 1912–1920.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288–295.

Cancer Genome Atlas Research Network (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N. Engl. J. Med. *368*, 2059–2074.

Curk, T., Demsar, J., Xu, Q., Leban, G., Petrovic, U., Bratko, I., Shaulsky, G., and Zupan, B. (2005). Microarray data mining with visual programming. Bioinformatics *21*, 396–398.

FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A.R.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J.L., Lassmann, T., Itoh, M., Summers, K.M., Suzuki, H., et al. (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307–315.

Haferlach, T., Kohlmann, A., Wieczorek, L., Basso, G., Kronnie, G.T., Béné, M.-C., De Vos, J., Hernández, J.M., Hofmann, W.-K., Mills, K.I., et al. (2010). Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. J. Clin. Oncol. *28*, 2529–2537.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Kühn, R., Schwenk, F., Aguet, M., and Rajewsky, K. (1995). Inducible gene targeting in mice. Science *269*, 1427–1429.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Pasini, D., Cloos, P.A.C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb

repressive complex 2 to target genes in ES cells. Nature *464*, 306–310.

Pedersen, M.T., Agger, K., Laugesen, A., Johansen, J.V., Cloos, P.A.C., Christensen, J., and Helin, K. (2014). The demethylase JMJD2C localizes to H3K4me3-positive transcription start sites and is dispensable for embryonic development. Mol. Cell. Biol. *34*, 1031–1045.

Pronk, C.J.H., Rossi, D.J., Månsson, R., Attema, J.L., Norddahl, G.L., Chan, C.K.F., Sigvardsson, M., Weissman, I.L., and Bryder, D. (2007). Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. Cell Stem Cell *1*, 428–442.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Quivoron, C., Couronné, L., Valle, Della, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.-H., et al. (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell *20*, 25–38.

Rapin, N., Bagger, F.O., Jendholm, J., Mora-Jensen, H., Krogh, A., Kohlmann, A., Thiede, C., Borregaard, N., Bullinger, L., Winther, O., et al. (2014). Comparing cancer vs normal gene expression profiles identifies new disease entities and common transcriptional programs in AML patients. Blood *123*, 894–904.

Ritchie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A., and Smyth, G.K. (2007). A comparison of background correction methods for two-colour microarrays. Bioinformatics *23*, 2700–2707.

Smyth, G.K., and Speed, T. (2003). Normalization of cDNA microarray data. Methods *31*, 265–273.

Somervaille, T.C.P., and Cleary, M.L. (2006). Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. Cancer Cell *10*, 257–268.

Ventura, A., Kirsch, D.G., McLaughlin, M.E., Tuveson, D.A., Grimm, J., Lintault, L., Newman, J., Reczek, E.E., Weissleder, R., and Jacks, T. (2007). Restoration of p53 function leads to tumour regression in vivo. Nature *445*, 661–665.

Wang, J., Wen, S., Symmans, W.F., Pusztai, L., and Coombes, K.R. (2009). The bimodality index: a criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data. Cancer Inform *7*, 199–216.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343–348.

Yan, M., Kanbe, E., Peterson, L.F., Boyapati, A., Miao, Y., Wang, Y., Chen, I.-M., Chen, Z., Rowley, J.D., Willman, C.L., et al. (2006). A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. Nat Med *12*, 945–949.

Ye, T., Krebs, A.R., Choukrallah, M.-A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res. *39*, e35.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.