Supplemental Information

TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells

Kasper D. Rasmussen, Ivan Berest, Sandra Kessler, Koutarou Nishimura, Lucía Simón-Carrasco, George S. Vassiliou, Marianne T. Pedersen, Jesper Christensen, Judith B. Zaugg and Kristian Helin.

Table of content

Figure S1. Rasmussen *et al*

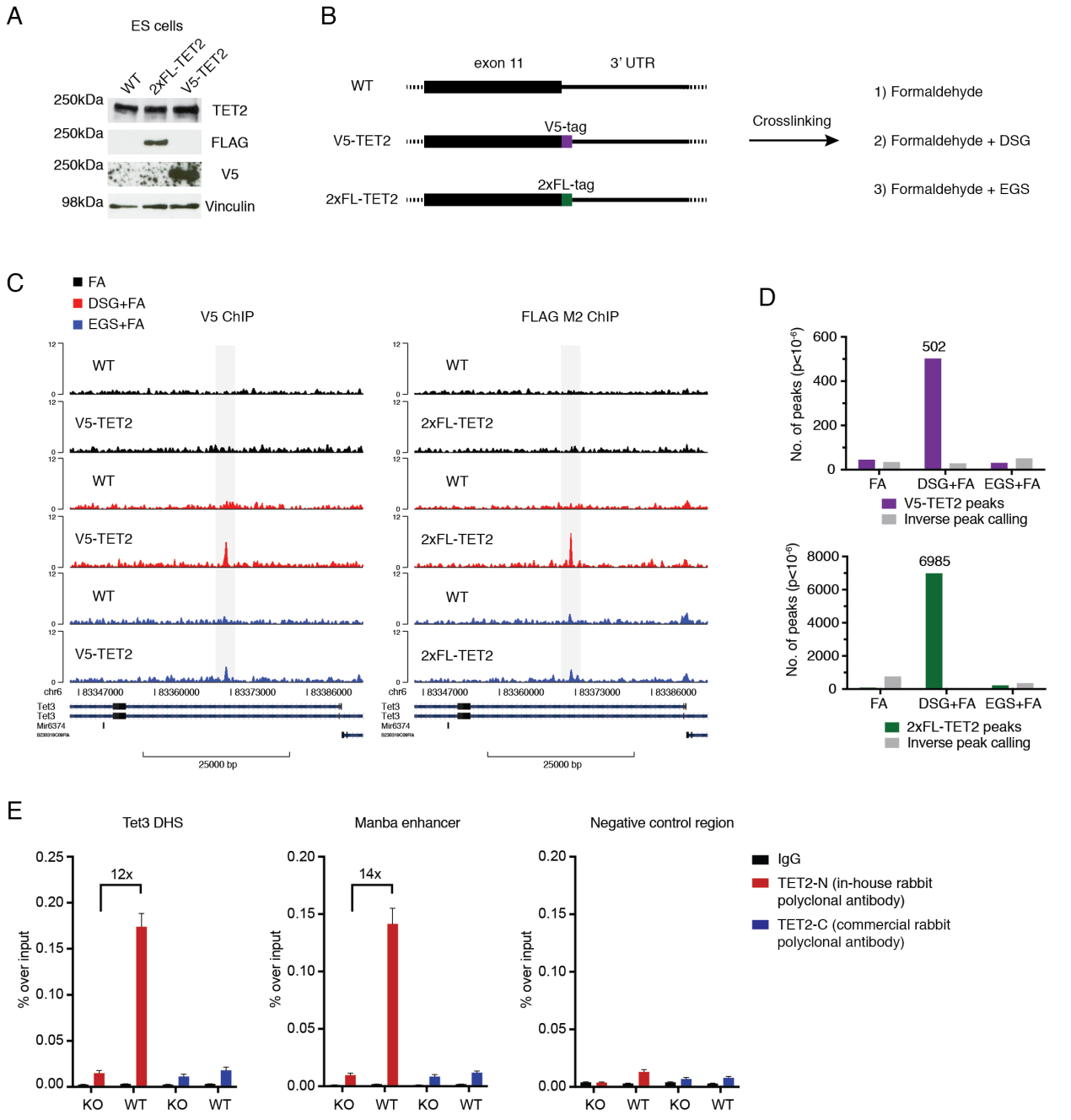


Figure S1: TET2 chromatin immunoprecipitation and sequencing on wildtype and epitope-tagged cell lines. Related to Figure 1

(A) Western blot for endogenously tagged TET2 in ES cells. Previous studies, in our lab and others, have been hampered by ChIP artifacts (recently coined as "Phantom peaks" (Jain et al. 2015)) and the lack of ChIP-grade antibodies against TET2. To circumvent this, we used CRISPR homology directed repair to introduce ChIP-verified epitope tags into the endogenous *Tet2* locus in embryonic stem (ES) cells. We identified, upon clonal expansion and characterization, two independent lines with in-frame integration of a V5-tag or a 2xFLAG (2xFL)-tag at the C-terminus of *Tet2.*

(B) Schematic showing cell lines expressing endogenously tagged TET2 and crosslinking conditions used for ChIP-seq.

(C) Representative ChIP-seq tracks of either V5 (left panel) or 2xFL (right panel) ChIP experiments showing TET2 binding site in the intron of *TET3* in ES cells. Analysis of ChIP coupled with high-throughput sequencing (ChIP-seq) using V5 or FLAG-specific antibodies revealed that the presence of the protein-protein crosslinker disuccinimidyl glutarate (DSG) in combination with formaldehyde (FA) greatly increased the signal-to-noise ratio of the ChIP signal and increased the number of called peaks. In contrast, crosslinking conditions with formaldehyde alone, or the combination of the longer ethylene glycol bis-(succinimidyl succinate) (EGS) crosslinker plus formaldehyde, were insufficient to capture these regions

(D) The number of called TET2 peaks are increased when crosslinking chromatin with DSG+FA. Histograms showing the number of called peaks in the different crosslinking conditions in V5-TET2 (upper panel) or 2xFL-TET2 (lower panel)-expressing cells compared to parental cells without endogenously tagged Tet2. The number of WT-specific peaks are also shown (inverse peak calling).
(E) ChIP-qPCR for TET2 binding in ES cells targeting two positive regions identified in (C and D) as well as a negative control region. Chromatin was crosslinked with DSG+FA and immunoprecipitation was performed using the rabbit polyclonal TET2-N antibody (Helin lab) or a rabbit polyclonal antibody directed against the Tet2 C-terminus (commercially available). We observed a strong enrichment of TET2-bound regions using a polyclonal rabbit antibody raised against the N-terminus of TET2 (TET2-N), and this signal was absent in knockout cells or at a negative control region. Thus, the combination of specific antibodies and optimized workflows allows for the precise mapping of TET2-bound genomic regions. Data is presented as mean enrichment over input with error bars depicting technical triplicates. Fold enrichment over *Tet2* knockout control cells is indicated.

Figure S2. Rasmussen *et al*

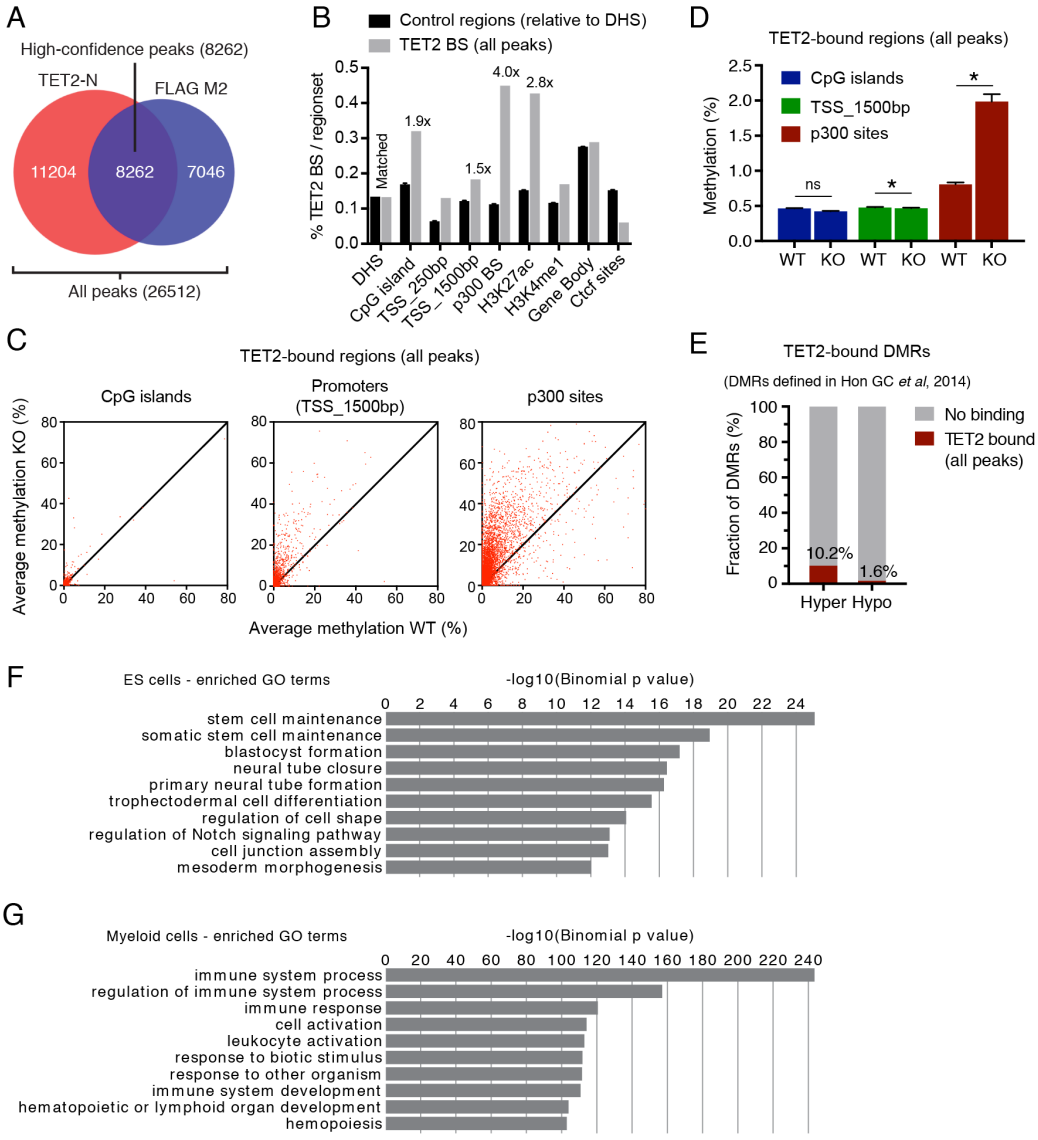


Figure S2: Analysis of TET2 binding sites in ES and hematopoietic cells. Related to Figure 1 and 2.
(A) Venn diagram as in Fig. 1A, but showing the origin of the low-stringency *all* peaks dataset in ES cells.
(B) Histogram showing overlap of genomic regions as in Fig. 1E, but for all TET2 peaks. Within this peak set, TET2 binding is enriched in p300 and H3K27ac regions, in agreement with high-confidence TET2 binding sites, and also show modest enrichment at CpG islands and promoters (1.9x and 1.5x, respectively).
(C) Scatter plots showing DNA methylation (5mC) changes in CpG islands (left), promoters (middle), and p300 sites (right). Each dot represents the average 5mC value at a region that overlaps with a TET2 binding site (all peaks). Only CpG sites covered by >10 reads were included in the analysis.
(D) Histogram showing data as in (C), but plotting the cumulative DNA methylation change of all regions in CpG islands, promoters or p300 binding sites. Data is presented as median +/- 95% confidence intervals. *, $P < 0.0001$ (paired two-tailed Student's t-test).
(E) Only a fraction of previously defined differentially methylated regions (DMRs) in *Tet2* knockout ES cells show evidence of TET2 binding. Histogram showing overlap of TET2 binding sites (all peaks) with DMRs (Hyper- or hypomethylated) defined by (Hon GC *et al*, 2014).
(F) Histogram showing enriched gene ontology (GO) terms identified by GREAT (McLean et al. 2010) analysis using promoter-distal TET2 binding sites (TET2-N only) from ES cells.
(G) Same as (F), but for myeloid hematopoietic cells.

Figure S3. Rasmussen *et al*

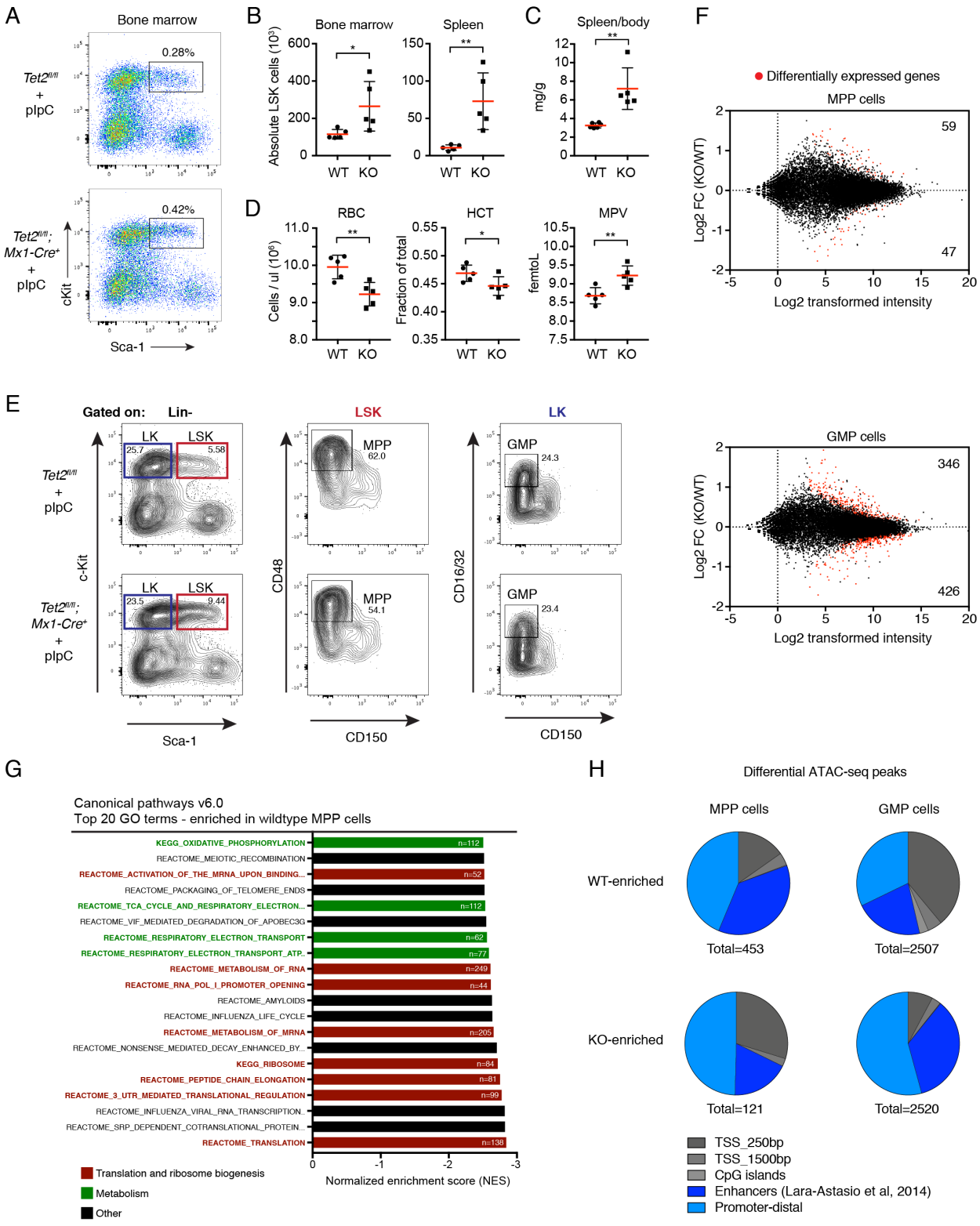


Figure S3: Phenotypic and genomic profiles of *Tet2*-deficient native hematopoiesis. Related to Figure 3.
(A) Representative FACS plots showing the Lineage-negative, cKit-positive, Sca-1-positive (LSK) population in the bone marrow of aged animals with wildtype or *Tet2*-deficient hematopoiesis. The percentages of total live bone marrow cells in the LSK gate are shown (representative of n=5).
(B) FACS enumeration of the absolute number of LSK cells in bone marrow (left) or spleen (right) in the wildtype or *Tet2*-deficient hematopoietic system of aged animals. Line and error bars represent mean +/- SD (n=5). *, $P < 0.05$. **, $P < 0.01$ (Unpaired two-tailed Student's t-test).

(C) Scatter dot plot showing the ratio of spleen and body weights of aged animals with wildtype or *Tet2*-deficient hematopoiesis. Line and error bars represent mean +/- SD (n=5). **, P < 0.005 (Unpaired two-tailed Student's t-test).
(D) Plots showing measurements of red blood cells (RBC), hematocrit value (HCT) and mean platelet volume (MPV) in aged animals with wildtype or *Tet2*-deficient hematopoiesis. Line and error bars represent mean +/- SD (n=5) (unpaired two-tailed Student's t-test).
(E) Representative FACS plots showing MPP and GMP sorting gates and the corresponding staining pattern of hematopoietic progenitors in aged animals with wildtype and *Tet2*-deficient hematopoiesis. The percentages with respect to the parent cell gate is shown for each population.
(F) Differential gene expression in wildtype and *Tet2*-deficient MPPs (upper) and GMPs (lower). Red dots indicate significant gene expression changes (FDR < 0.1) and the sum of differentially expressed genes are indicated in the corresponding quadrants.
(G) GSEA results showing top 20 gene ontology (GO) terms enriched in wildtype compared to *Tet2*-deficient MPPs. Gene ontology terms related to translation, ribosome biogenesis and metabolism are marked in red and green, respectively. The number of genes in each signature is indicated.
H) Pie charts showing the genomic distribution of differential regions of open chromatin from ATAC-seq analysis of wildtype or *Tet2*-deficient MPPs and GMPs. Blood lineage enhancers defined in (Lara-Astiaso et al. 2014) were used for overlap analysis.

Figure S4. Rasmussen *et al*

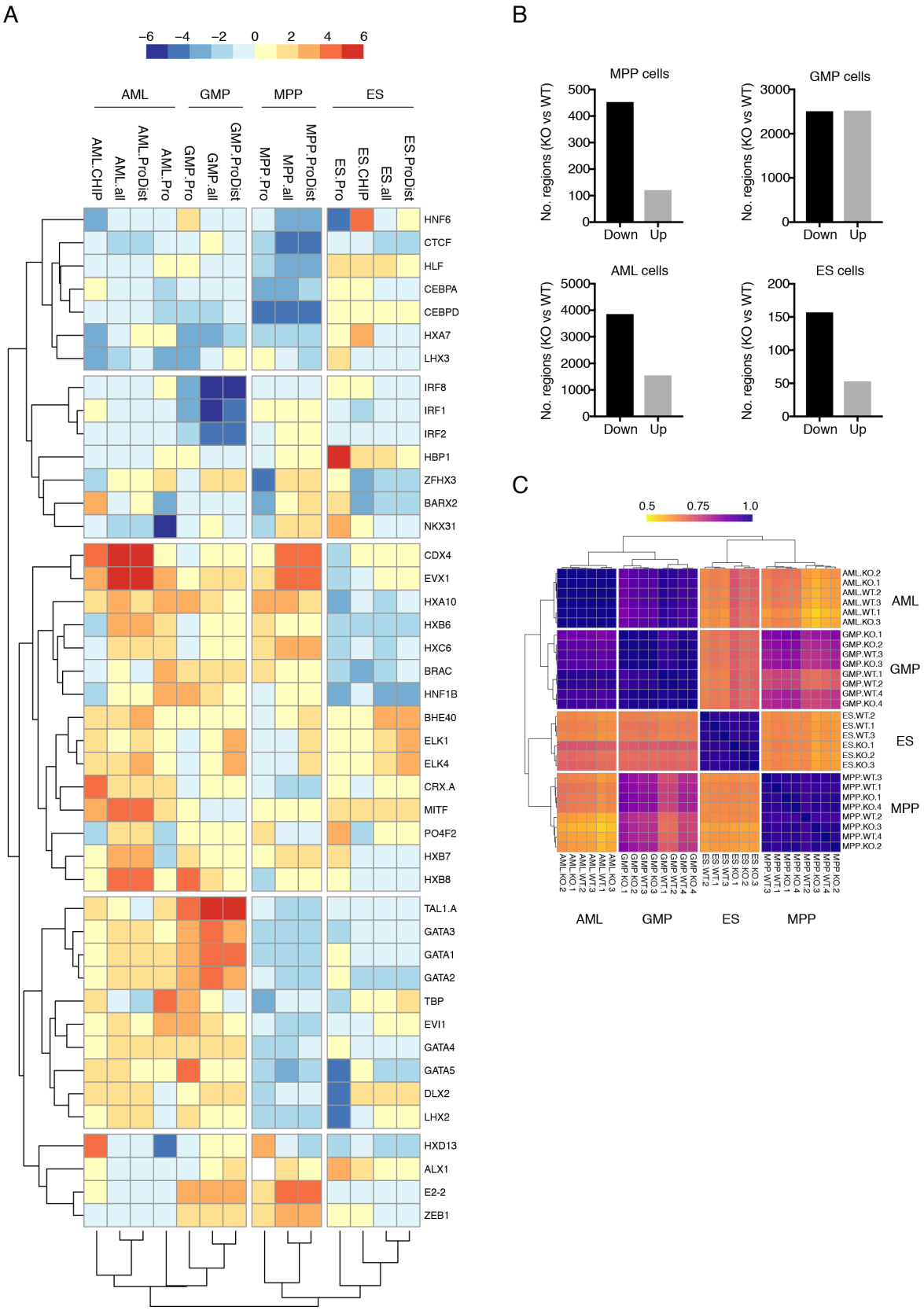


Figure S4: Summary of *diffTF* analysis and ATAC-seq profiles of multiple *Tet2*-deficient cell types. Related to Figure 4 and 5.

(A) Summary of differential TF activity between wildtype and *Tet2* knockout across multiple cell types. The heat map represents the summarized *diffTF* output for TET2-WT versus TET2-KO analyses in AML, GMP, MPP and ESC. For each cell type the analysis was run on different sets of peaks: promoter regions only (-1.5 kb/+500bp from TSS; labeled

celltype.Pro) and putative enhancer regions (+100kb/-100kb from TSS excluding promoter regions; *celltype*.ProDist), in addition to the full set of peaks (*celltype*.all). For the cell types for which we mapped TET2 binding sites by ChIP-Seq (ESC and AML) *diffTF* was also run for the ATAC-seq peaks intersected with TET2 ChIP in corresponding cell type (*celltype*.CHIP). TET2 bound regions in AML cells were determined in bulk AML cells (containing both the leukemic precursor population that was used for ATAC-seq analysis as well as more differentiated cells). The color displays z-scores of weighted mean difference values obtained from diffTF. TFs with the largest signal across cell types are displayed (0.1 and 0.9 quantiles). Positive values indicate that the TF has higher activity in TET2 WT samples, and vice versa for negative values.

(B) The number of regions with significantly altered chromatin accessibility (FDR < 0.05) identified by *diffBind* analysis of replicate ATAC-seq profiles of wildtype and *Tet2*-deficient MPP cells (n=4), GMP cells (n=4), AML cells (n=3), and ES cells (n=3).

(C) Heat map showing the overall Pearson correlation of ATAC-seq profiles of the four cell types. AML cells show the strongest similarity to GMP cells, whereas ES cells are distinct from the hematopoietic cell populations.

Figure S5. Rasmussen *et al*

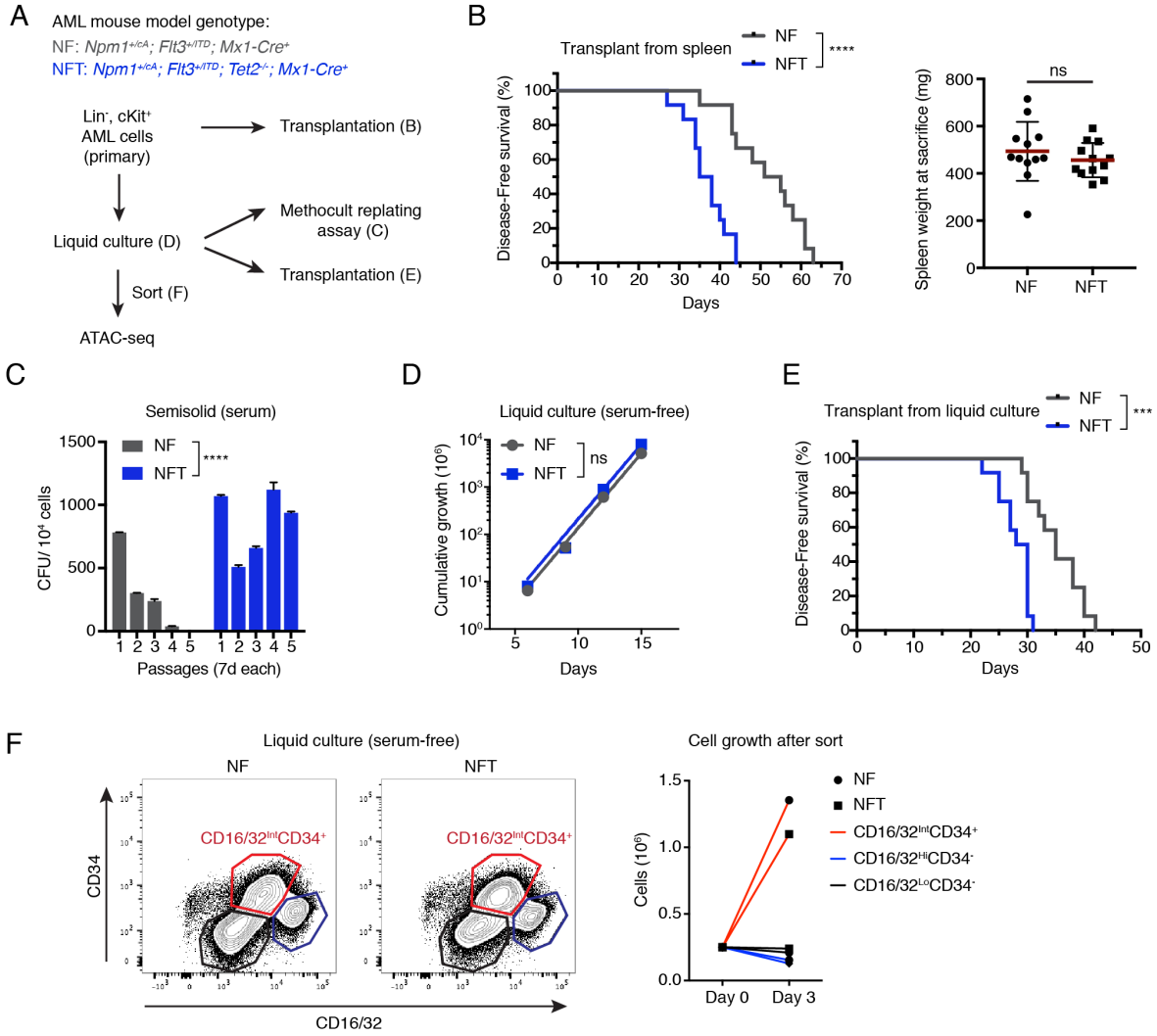


Figure S5: Generation and analysis of a novel mouse model of human AML with mutations in *TET2* as well as the frequently co-occurring mutations *NPM1c* and *FLT3-ITD*. Related to Figure 5.

(A) Diagram showing experimental setup to analyse AML cells with and without deletion of *Tet2*. Genetically engineered mouse models (Vassiliou et al. 2011; Lee et al. 2007; Quivoron et al. 2011) carrying the indicated alleles were crossed and monitored for disease development. As previously described (Velasco-Hernandez et al. 2016), leaky Cre-activity induces AML formation in the absence of pIpC-mediated induction of Cre recombinase. Liquid cultures in serum-free conditions were started by sorting and plating ckit^{+}Gr1^{-}Mac^{-} cells from primary leukemic splenocytes. These cells were subsequently used for *in vivo* and *in vitro* analysis, and CD34^{+}CD16/32^{int}CD11b^{-}Gr1^{-} cells were sorted for analysis of chromatin accessibility by ATAC-seq.

(B) Kaplan-Meier plot (left panel) showing disease-free survival and spleen weight at sacrifice (right panel) of recipient mice transplanted with AML cells. Sublethally irradiated (650 Rad) Ly5.1 mice (n=12) were each transplanted with 2.5 x 10^4 sorted c-Kit^{+}Gr1^{-}Mac1^{-} cells isolated from spleens of moribond animals with the indicated genotypes. ns, not significant. ****, $P < 0.0001$ (Log-rank (Mantel-Cox) test).

(C) Histogram showing colony-forming activity (CFU) of AML cells from liquid culture with the indicated genotypes. The cells were grown in semi-solid methylcellulose-containing media (Methocult GF M3435) and colonies were counted at each replating (7 days each). The experiment was repeated twice with the same result. The data represents mean +/-SD of technical duplicates. ****, $P < 0.0001$ (2-way ANOVA).

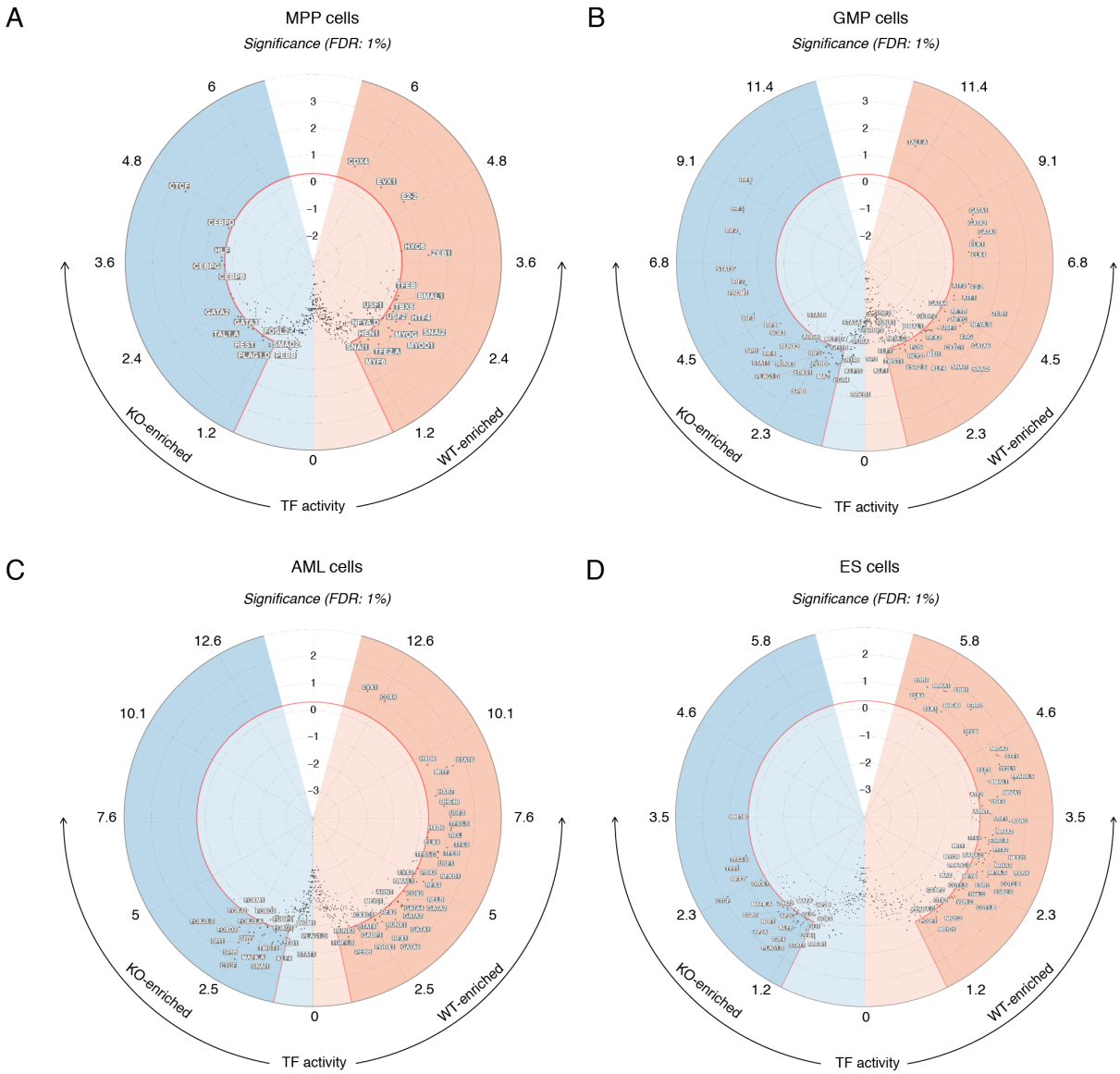
(D) Cumulative cell numbers in liquid cultures of AML cells grown in serum-free conditions. There is no significant difference in accumulation of cells. ns, not significant (Non-linear regression).

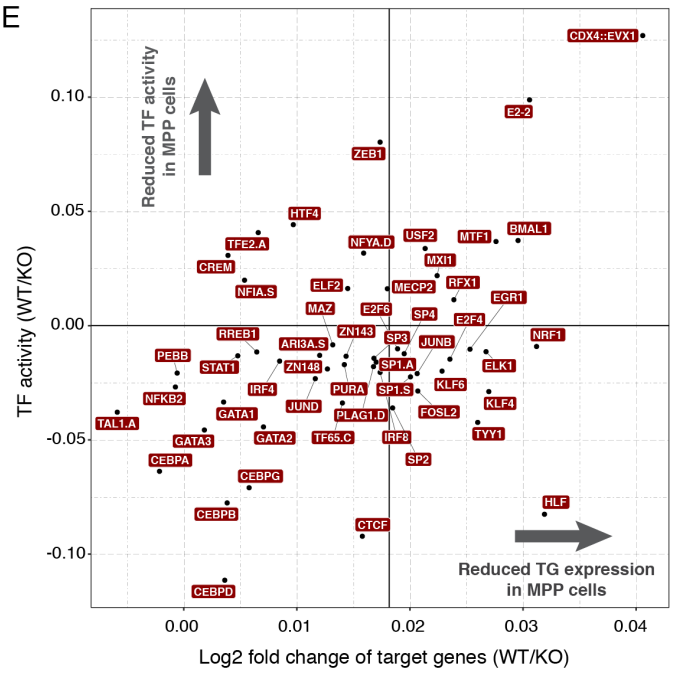
(E) same as (B), but for sublethally irradiated Ly5.1 mice transplanted with 1 x 10^6 bulk AML cells from liquid culture. ***, $P = 0.0001$ (Log-rank (Mantel-Cox) test).

(F) Representative FACS plots (left panel) of AML cells from liquid culture stained for CD34 and CD16/32 expression. The three indicated cell populations were sorted and allowed to re-establish the liquid culture (right panel). Only the

cell population expressing CD34 (and with low Gr1 and Mac1 expression) was able to grow and reform the original cellular hierarchy of AML cells in the culture.

Figure S6. Rasmussen *et al*





F

ES cells - top enriched motifs in CDX4 BS

Motif found	Known or similar motif	E-value
	CDX1 CDX2	1.1e-529
	KLF5 SP1 ASCL2	1.5e-218
	SOX15 SOX2 SOX9	9.4e-185
	POU3F1 POU2F3	1.5e-158

Figure S6: Overview of *diffTF* results and correlation with target gene expression. Related to Figure 4, 5 and 6.
A-D. Differential TF activity analyses for TET2-WT versus TET2-KO in MPP (A), GMP (B), AML (C), and ES (D) cells. The circular volcano plot represents the *diffTF* analysis results from promoter-distal (putative enhancers) consensus peaks for the comparison of WT vs TET2 KO in the indicated cell type. Some TFs are significantly more active in TET2-WT (labelled TFs in the orange background), and TET2-KO (labelled TFs in the blue background). The radial (y-axis) shows the –log10 p-value and the red line indicates the threshold for an FDR of 0.01. The threshold on the x-axis is defined by the 0.1 and 0.9 quantiles of the data.
(E) Correlation of changes in TF activity and target gene expression. The scatter plot shows a positive correlation (Pearson's r=0.4; p-value < 0.005) between TF activity, as defined by *diffTF*, and the median log2 fold change of putative target genes of the same TF. Target genes for each TF were defined as genes that had a TF binding site within +/-100kb of the promoter region (using the *annotatePeak* function in "ChIPseeker" Bioconductor package). The median log2 fold change of the target genes expression was calculated based on all genes that had a mean of expression of more than 10 counts. Only expressed TFs that had more than 50 target genes and an FDR > 0.05 in the *diffTF* analysis are shown. The accentuated vertical line (x = 0.0175) represents the median expression of all target genes used in the analysis.
(F) List of top enriched sequence logos and their associated DNA-binding TFs identified in CDX4-bound regions in ES cells by MEME-ChIP (Machanick and Bailey 2011).

Supplemental Experimental Procedures

ES cell lines

Mouse ES cell lines were derived from blastocysts with the following genetic background *Tet2*$^{fl/fl}$ (exon 3 deletion allele (Moran-Crusio et al. 2011)) or *Tet2*$^{fl/fl}$*; Rosa26*$^{+/Cre\text{-}ERT2}$ (exon 11 deletion allele (Quivoron et al. 2011)). To generate a clean TET2 knockout ES cell line for ChIP control (Fig. 1), the parental *Tet2*$^{fl/fl}$ ES cell line (derived from (Moran-Crusio et al. 2011)) was transiently transfected with a Cre recombinase-expressing plasmid and subcloned to identify a constitutive *Tet2* knockout ES cell line. All mouse ES cell lines were cultured in feeder-free, gelatinized plates in "Serum/2i/LIF"-conditions as described in (Hon et al. 2014).

Endogenous tagging of TET2 was performed using CRISPR homology directed repair with a single-stranded DNA oligonucleotide repair template. Briefly, a sgRNA targeting the insertion site was cloned into pSpCas9(BB)-2A-GFP (PX458), deposited by Feng Zhang (Addgene #48138). This plasmid was co-transfected in ES cells with oligonucleotides encoding either two copies of Flag (DYKDDDDKDYKDDDDK) or a single copy of V5 tag (GKPIPNPLLGLDST) as well as 60 nucleotide homology arms. Transfected cells were single-cell sorted and the resulting clones screened for stable expression of epitope-tagged TET2 by western blot.

Stable expression of epitope-tagged CDX4 was performed using PiggyBAC (PB) transgenesis in ES cells. Briefly, mouse *Cdx4* cDNA with two copies of FLAG tag at the amino terminus was amplified by PCR and cloned into a PB vector encoding a blasticin resistance gene (*PB-IRES-Bsd)*. This vector was co-transfected with pBASE into *Tet2*$^{fl/fl}$*; Rosa26*$^{+/Cre\text{-}ERT2}$ ES cells and selection with 5µg/ml blasticidin was started 24 hours post-transfection. After 7 days in selection media, single cells were isolated by limited dilution and the resulting clones screened for stable expression of 2xFL-CDX4 by western blot. To activate Cre-ERT2, the ES cell media was supplemented with 500nM 4-Hydroxytamoxifen (4-OHT).

Mice

For analysis of aging-related *Tet2*-deficient hematopoiesis, cohorts of age-matched litter mates (8-weeks old) of *Tet2*$^{fl/fl}$ or *Tet2*$^{fl/fl}$*: Mx1-Cre*$^{+}$ mice (Quivoron et al. 2011) were injected 3 times intraperitoneally with

250µg polyinosinic-polycytidylic acid (PolyI:C LMW, InvivoGen) dissolved in saline at experimental days 0, 2, and 4. The mice were subsequently allowed to age and sacrificed at 10 months of age. To generate *Tet2*-deficient AML model, the following genetically modified mouse lines *Npm1*$^{cA\text{-}Flox}$ (Vassiliou et al. 2011), *Flt3-ITD* (Lee et al. 2007), *Tet2*$^{fl/fl}$; *Mx1-Cre+* (Quivoron et al. 2011) were intercrossed and *Npm1*$^{+/cA\text{-}Flox}$*; Fl3*$^{+/ITD}$*; Mx1-Cre*$^{+}$ or *Npm1*$^{+/cA}$*; Fl3*$^{+/ITD}$*; Tet2*$^{fl/fl}$*; Mx1-Cre*$^{+}$ mice were monitored for disease development by taking advantage of leaky Mx1-Cre activity in the presence of the Flt3-ITD allele (Velasco-Hernandez et al. 2016). AML cell were harvested and successful recombination of loxP-flanked alleles (*Npm1*$^{cA\text{-}flox}$ and *Tet2*$^{fl/fl}$) were verified by PCR. For secondary transplantations, sublethally irradiated (650Rad) Ly5.1 recipient animals were transplanted by tail vein injection with sorted c-Kit^{+}Gr1^{-}Mac1^{-} AML splenocytes harvested from moribound mice (25,000 cells/mouse) or after 4-6 weeks of *in vitro* cell culture (1*10^{6} bulk in vitro cells/mouse). Ly5.1 mice were maintained on medicated water (Ciprofloxacin 100µg/ml) for three weeks following the irradiation procedure. All animal work was carried out in compliance with ethical regulation under license by the Danish regulatory authority.

***In vitro* culture of AML cells**

To establish *in vitro* culture of AML cells, c-Kit^{+}Gr1^{-}Mac1^{-} splenocytes harvested from moribund mice were purified by FACS and cultured in suspension in non-tissue culture treated plasticware in StemPro-34 SFM media (ThermoFisher Scientific, Cat: 10639011) supplemented with 2mM GlutaMAX (GIBCO), 1X pen/strep (GIBCO), 0.1mM 2-mercaptoethanol (SIGMA), as well as the cytokines SCF (50ng/ml), IL-3 (10ng/ml), and IL-6 (10ng/ml) (PEPROTECH). To assess colony-forming unit activity, stably growing AML cells were plated in triplicate in methylcellulose medium (M3534, Stem Cell technologies) according to manufacturer's protocol. Cells were seeded with equal density (10,000 cells/plate) every 7 days and colonies were counted at each replating.

Lentiviral transduction

Lentiviral vectors for expression of wildtype and catalytic-inactive TET2 c-terminal domain as well as empty control were gifts from Silvia Monticelli (Addgene plasmids #79611, #79554, #99636). 293FT cells were

transfected with lentiviral vectors and packaging plasmids psPAX2 and pMD2.G using a standard calcium phosphate protocol. The supernatant was collected 72 hours after transfection and viral particles were concentrated by ultracentrifugation at 24,000 x g for 2 hours. The transduced AML cells were selected with 2 ug/ml puromycin 24 hours after transduction and robust expression of TET2 C-terminal fragments were verified by western blotting.

Fluorescence-activated cell sorting (FACS)

Single-cell suspensions of mouse bone marrow were erythrolysed, enriched for Kit expression (CD117 microbeads, Miltenyi Biotech) and stained with antibodies against surface markers: Lineage (B220-PECy5 (RA3-6B2, eBioscience), CD11b-PECy5 (M1/70, eBioscience), Ter119, PECy5 (TER-119, eBioscience), CD3e-PECy5 (145-2C11, eBioscience), Gr1-PECy5 (RB6-8C5, eBioscience)), Sca1-BV421 (D7, BD biosciences), cKit-AlexaFlour 780 (2B8, eBioscience), CD150-APC (TC15-12F12.2, Biolegend), CD48-PE (HM48-1, eBiocience), CD16/32-PECy7 (93, eBioscience). The following combination of surface markers was used to define hematopoietic progenitor populations: Multipotent Progenitors (MPPs), $Lin^{-}cKit^{+}Sca1^{+}CD48^{+}CD150^{-}$; Granulocyte-Monocyte Progenitors (GMPs), $Lin^{-}cKit^{+}Sca1^{-}CD150^{-}CD16/32^{+}$.

AML cells cultured *in vitro* were harvested and stained with antibodies against the surface markers: CD11b-PE (M1/70, eBioscience), Gr1-AlexaFlour 700 (RB6-8C5, eBioscience), CD16/32-PECy7 (93, eBioscience), CD34-FITC (RAM34, eBioscience). The following combination of surface markers was used to define the leukemic precursor population purified for ATAC-seq analysis: $CD11b^{-}Gr1^{-}CD16/32^{int}CD34^{+}$. Total live bone marrow cells were stained with CD317-FITC (PDCA-1, eBioscience) and B220-APC (RA3-6B2, eBioscience) to enumerated plasmacytoid dendritic cells (pDCs). All cells populations were purified on FACSAria III (BD Biosciences) and flow cytometry data was analysed using FlowJo software (Tree Star inc.).

Western blotting

Cells were lysed by boiling in 2x LBS sample buffer and Western blotting was carried out according to standard protocols with the primary antibodies to CDX4 (Abcam, ab133145), TET2 (rabbit polyclonal antibody, Helin lab), Vinculin (Sigma, F7425), V5-epitope (Abcam, ab15828), FLAG-epitope (Sigma, F7425),

and Myc-epitope (9E10). Bands were visualized using Super Signal West Pico chemiluminescent ECL substrate (ThermoFisher Scientific) for exposing on Amersham hyperfilm ECL films (GE healthcare).

Quantitative real-time PCR

Total RNA was extracted from FACS purified hematopoietic cell populations using RNeasy Micro Kit (Qiagen). Quality assessment of total RNA was performed using an RNA 6000 Pico Kit (Agilent) and only samples with a RNA integrity number above 8 were used for further analysis. RNA concentrations were determined using a Qubit RNA HS assay kit (ThermoFisher Scientific) and a Qubit 2.0 fluorometer. Reverse transcription was performed with TaqMan Reverse Transcription reagents kit with random hexamers (Applied Biosystems). Real-time PCR was carried out using 2X SYBR Green I master mix (Roche) on LightCycler 480 Real-Time PCR system (Roche) with primers to *Cdx4* (Forward: CTGGGCCTTTCTGAGAGACA, Reverse: GTCACTTTGCACGGAACCTC), *Evx1* (Forward: AGAAATCGAGGTGAGCTGCA, Reverse: GCCACTGGTAGGACTGCTAC), and *Hprt* (Forward: CTGGTGAAAAGGACCTCTCG, Reverse: CAAGGGCATATCCAACAACA). If biological replicate samples yielded a late Cp call (> 40 cycles) the gene in question was defined as "not detected".

RNA-seq library generation and analysis

A proportion of total RNA (2 ng) isolated from four biological replicates (individual mice with wildtype or *Tet2*-deficient hematopoiesis) was amplified and size-selected with the Ovation RNA amplification system v2 (NuGen, Cat: 7102) and sequencing adaptors were added to the resulting cDNA using the Ovation Ultra Low v2 system (NuGen) according to manufacturer´s instructions. RNA-seq libraries were quality checked using a DNA 1000 Kit (Agilent) and sequenced using an Illumina NextSeq 550 instrument (75bp single-end).

Raw sequencing reads were adaptor trimmed and aligned to the mouse genome (mm10) with STAR aligner using default parameters. Mapped reads were assigned to genes and differential gene expression was analysed using DESeq2. Four biological replicates were used for the DESeq2 analysis. Gene counts were normalized by variance stabilizing transformation (VST) and *z*-score plotted in the heat map for genes with q-value < 0.1 in both GMP and MPP cells (see Table S1 for full list of deregulated genes). Gene set

enrichment analysis (GSEA) was performed on VST gene counts against curated gene sets of canonical pathways (c2.cp.v6.0.symbols.gmt) in the MSigDB database (software.broadinstitute.org). Statistical significance of GSEA results were determined by gene set permutation (10^4 permutations) (see Table S2 for full list of GO terms).

ChIP-seq library generation

ES cells or hematopoietic cells were washed twice in ice-cold PBS and pelleted by centrifugation. The cells were resuspended in 10 ml ice-cold PBS and freshly prepared 0.25M disuccinimidyl glutarate (DSG) stock solution (dissolved in DMSO) to obtain a final concentration of 2mM DSG in PBS (ThermoFisher Scientific, Cat: 20593) and incubated at a rotating wheel for 30 min to allow equilibration to room temperature. Then, formaldehyde (Sigma, Cat: 252549) was added to obtain a final concentration of 1% and rotated for another 10 min at room temperature. Finally, the crosslinking reaction was stopped by addition of glycine to a final concentration of 125mM. The cells were spun down for 5 min at 350 x g rpm at room temperature and washed twice with ice-cold PBS. The cells were then resuspended in 5 mL SDS Buffer (50mM Tris-HCl pH 8.1, 100mM NaCl, 5mM EDTA, 0.5% SDS) containing 1 mM Phenylmethylsulfonyl fluoride (PMSF) and allowed to rotate for another 5 min. Finally, the cells were pelleted and resuspended in IP buffer (100mM Tris-HCl pH 8.6, 100mM NaCl, 5mM EDTA, 0.3% SDS, 1.7% TritonX-100) with proteinase inhibitors according to pellet size. Chromatin dissolved in IP buffer was sheared to an average size of 200-500 bp DNA fragments in a Bioruptor (Diagenode). The sonicated chromatin was diluted in SDS-free IP buffer to achieve a concentration of 0.1% SDS, spun down at 20,000 x g for 20 min to remove insoluble chromatin fraction and precleared with protein G Sepharose beads (GE healthcare) prior to immunoprecipitation.

For immunoprecipitation of endogenous TET2, 1µg affinity-purified rabbit polyclonal antibody raised against N-terminal TET2 protein (TET2-N) (Helin lab) was incubated with 300µg chromatin overnight. The chromatin-antibody complexes were captured in a 3h incubation with protein-G Sepharose beads (GE healthcare). For immunoprecipitation of 2xFL-TET2 or 2xFL-CDX4, 20µl of anti-FLAG M2 affinity gel (Sigma, A2220) was incubated with 300ug chromatin for 3h. Washes of chromatin-antibody-bead complexes were performed as follows: Three washes with ice-cold 150mM wash buffer (20mM Tris-HCl pH 8.0, 150mM

NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100), two washes with ice-cold 500mM wash buffer (20mM Tris-HCl pH 8.0, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100), and one wash with ice-cold IP buffer with a final concentration of 0.1% SDS. After the last wash, DNA from TET2-N immunoprecipitations was decrosslinked by overnight incubation at 65^{o}C in decrosslinking solution (1% SDS, 0.1M $NaHCO_3$). In FLAG M2 immunoprecipitations, IP'ed chromatin was initially eluted by three consecutive incubations (each 20 min on ice) in elution buffer (20mM Tris-HCl pH 8.0, 150mM NaCl, 2mM EDTA) with 0.5mg/ml FLAG peptide (Peptide 2.0). The eluted fractions were pooled and decrosslinked by overnight incubation at 65^{o}C in decrosslinking buffer. IP'ed DNA was purified using the QIAquick PCR purification kit (Qiagen) according to manufacturer's instructions.

ChIP-seq libraries for Illumina sequencing was prepared using the NEBNext Ultra II DNA library preparation kit (New England Biolabs) using an input of 1-3ng of IP'ed DNA (quantified using DNA HS assay kit (Qubit)) following the manufacturer's instructions. Adaptor-ligated fragments were size-selected using AMPure XP beads (Beckman Coulter) to retain inserts of approximately 200bp prior to PCR amplification. Equimolar amounts of sample, with compatible indexes, were pooled and sequenced on Illumina NextSeq 550 (75bp single-end).

Processing and analysis of ChIP-seq data

Raw sequencing reads were trimmed to remove low quality nucleotides and adaptors with Trimmomatic using ILLUMINACLIP and mapped to the mouse genome (mm10) using Bowtie2 (Langmead and Salzberg 2012) with (--very-sensitive) preset settings. Mapped reads were furthermore de-duplicated using PicardTools v2.7.1 markDuplicates function. Read trimming, mapping and de-duplication were performed using an in-house ChIP-seq processing pipeline (Snakemake) or the public server of the Galaxy Project (Afgan et al. 2016).

In ES cells, TET2 binding sites were defined by differential binding affinity analysis between biological replicate ChIP-seq samples. Briefly, peaks in ChIP experiments using either TET2-N or FLAG M2 antibodies were called over IgG background using MACS2 (Zhang et al. 2008) with 0.05 q-value cutoff. These sets of peaks were merged to derive a consensus peak set using *"Diffbind"* package from

Bioconductor with following parameters: minOverlap = 1, filter = 5. Statistically significant differentially bound regions (FDR < 0.1 and log2 fold change > 0.5) between wildtype and *Tet2* knockout (or FLAG-tagged and empty) were identified by DESeq2 corresponding to 19,480 and 15,322 TET2 binding sites in TET2-N and FLAG M2 ChIP experiments, respectively.

Identification of sites with evidence of TET2 binding in hematopoietic cells was, due to lower signal-to-noise ratio, performed with less stringent criteria using MACS2 with default parameter settings and a q-value cutoff of 10^{-06}. Peaks were called based on mapped reads from paired samples of wildtype and *Tet2* knockout cells, and peak sets from biological duplicate experiments were concatenated and regions overlapping with at least 1 bp were merged. This resulted in identification of 19,706 TET2 binding sites in myeloid hematopoietic cells and 7,002 TET2 binding sites in bulk AML cells. CDX4 binding sites were defined as enriched regions in cells expressing 2xFL-CDX4 versus parental cells without 2xFL-CDX4 expression using MACS2 with default parameter settings and a q-value cutoff of 10^{-06}. This yielded 13,977 CDX4-bound regions (See Table S3 for genomic coordinates of all region sets used in the analysis).

Downstream analyses of mapped reads and visualization in heat maps, mean signal plots and as ChIP-seq tracks were performed using EaSeq 1.05 (Lerdrup et al. 2016). Random matched control regions for TET2 peak sets (with respect to gene bodies or DHS) were generated in EaSeq using the "Matched controls" tool. The following experimentally determined datasets were downloaded from the ENCODE project website (ENCODE Project Consortium 2011) for analysis of TET2 ChIP co-localization in functional genomic regions: DNase-seq "ENCFF001YOK" and "ENCFF00PUJ", p300 binding sites "ENCFF001YAD", H3K27 acetylation "ENCFF001XWY", H3K4 monomethylation "ENCFF001XWY", and CTCF binding sites "ENCFF001YAC". Active and poised enhancers in myeloid cells were defined as in (Rasmussen et al. 2015). DNA methylation data from wildtype and *Tet2* knockout ES cells were obtained from (Hon et al. 2014) (GEO: GSE48519). Motif enrichment analysis in ChIP-seq enriched regions were performed with MEME-ChIP (Machanick and Bailey 2011) against the HOCOMOCO v11 database.

ATAC-seq library generation

ATAC-seq libraries were generated as described previously (Buenrostro et al. 2013; Lara-Astiaso et al. 2014). Briefly, 10,000 freshly isolated cells (MPP, GMP, AML, and ES) were sorted by FACS into ice-cold FACS buffer (PBS + 2%FBS). The cells were pelleted using a swinging bucket centrifuge (500 x g, 10min, 4^{o}C) with settings for low acceleration/deceleration and washed once in ice-cold PBS. The cell pellets were resuspended in 50µl lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM $MgCl_2$, 0.1% Igepal CA-630) by gentle pipetting and immediately centrifuged one additional time (500 x g, 10min, 4^{o}C). The supernatant was discarded and the pellet containing released nuclei were resuspended gently in 25µl 1xTD buffer containing 1.25µl Tn5 transposase (Nextera sample preparation kit, Illumina). The transposition reaction was allowed to proceed for 45min at 37^{o}C whereafter DNA fragments were isolated using MinElute PCR purification columns (Qiagen) according to manufacturer's instruction.

To generate multiplex libraries, the transposed DNA was initially amplified for 5x PCR cycles using 2.5µl each of dual-index primers (Nextera index kit, Illumina) and 2.5µl PCR primer cocktail (PPC, Illumina) in a 25µl reaction volume of 1x KAPA HiFi hot-start ready-mix (Kapa BioSystems). The hot-start polymerase was activated prior to adding to the reaction mix by performing a brief pre-incubation step of 3min at 95^{o}C. The amplified fragments were size-selected with AMPure XP beads (0.5X) to remove fragments larger than 600bp and an aliquot was quantified to determine the optimal PCR cycle number to obtain 1/3 of maximum fluorescence intensity (Library quantification kit, Kapa Biosystems). Finally, PCR amplification was performed using the optimal number of cycles determined for each library (max. 18 cycles in total), size-selected with AMPure XP beads (0.5X) and eluted in resuspension buffer (Illumina). The size distribution of the libraries was evaluated on Bioanalyzer (Agilent) and sequenced on NextSeq 500 (Illumina) using 150bp or 75bp paired-end sequencing with an average of 25 million reads per sample.

Preprocessing of ATAC-seq data

We developed an in-house pipeline to preprocess the ATAC-Seq data, using commonly used tools and parameters. Briefly, it starts with raw fastq files then perform adapter trimming, alignment, and general and ATAC-Seq specific post-alignment filtering and processing steps. We used FastQC to assess the sequence quality, remove foreign sequences from the Nextera Transposase agent with Trimmomatic, using the parameters *ILLUMINACLIP:NexteraPE-PE.fa:1:30:4:1:true TRAILING:3 MINLEN:10*. Alignment was

performed with Bowtie2 with the default parameters *-X 2000* (maximal fragment length), *--very-sensitive* and against mm10, followed by various cleaning steps (Picard tools CleanSam, FixMateInformation, AddOrReplaceReadGroups, and ReorderSam).

We then removed mitochondrial reads, reads with a mapping quality (MAPQ) below 10, and duplicate reads with Picard tools. Finally, we adjusted the read start sites as described previously (4 bp on the forward and 5 bp on the reverse strand) and removed reads with insertions or deletions using samtools.

diffTF pipeline for analysis of TF activity

To assess differences in TF activity between conditions we developed an approach termed diffTF (Berest et al *in revision*). We defined "TF activity" as a change in chromatin accessibility (as measured by ATAC-Seq), near the putative binding sites of a TF. This assumption is based on the observation that many genetic variants that affect chromatin accessibility fall into TF binding sites, thus suggesting that it is a change in TF binding that affects the accessibility signal (Kumasaka et al. 2016).

To measure the differences of activity between two conditions we applied diffTF (see Berest et al *in revision).* Briefly, we first scanned the genome for TF binding sites using the mouse PWMs deposited in the database Hocomoco v10, and PWMScan to identify the binding sites for each TF. For each TF we then extracted the signal around each binding site (+/- 100bp around the core motif) and calculated a fold-change between two conditions. To account for differences in GC content at the different binding sites, we binned them into bins based on their GC content (0-10% - 90-100%), and for each bin, compare the distribution of fold-changes for binding sites of one TF to the background distribution of fold-changes across all other TFs to calculate the mean difference. To get one value per TF we then calculated the mean across all GC bins weighted by the number of its binding sites in each bin. To assess the significance of the differences we performed a Welch Two Sample t-test for each bin, and calculated the overall significance by treating the T-statistics as z-scores. This allowed to summarize them across all bins and convert them to one p-value per TF (see Berest et al *in revision*).

Correlation of diffTF values and target gene expression

Generally, we expect that the activity of a TF on chromatin, as we defined within the diffTF approach, should correlate with a TF's effect on target gene expression. To test whether this is indeed the case, and therefore to validate diffTF in this particular data set we sought to compare the effect of a TF on its target genes to its activity on chromatin. To do so, we obtained a set of target genes for each TF, calculated their median fold-change, and compared this value to the diffTF value of the same TF (Fig S6E). As a proxy, we defined target genes for each TF as those that had a TF binding site within +/-100kb of the promoter region, which should cover TFs that act in promoters as well as enhancers. We used the *annotatePeak* function in the "ChIPseeker" Bioconductor package to map TF binding sites to the +/-100kb windows of the genes. To calculate the fold-change across target genes, only genes that had a mean expression of at least 10 counts were taken into account.

Supplemental References

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**: W3–W10.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–1218.

ENCODE Project Consortium. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). ed. P.B. Becker. *PLoS Biol* **9**: e1001046.

Hon GC, Song C-X, Du T, Jin F, Selvaraj S, Lee AY, Yen C-A, Ye Z, Mao S-Q, Wang B-A, et al. 2014. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell* **56**: 286–297.

Jain D, Baldi S, Zabel A, Straub T, Becker PB. 2015. Active promoters give rise to false positive "Phantom Peaks" in ChIP-seq experiments. *Nucleic Acids Res* **43**: 6959–6968.

Kumasaka N, Knights AJ, Gaffney DJ. 2016. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat Genet* **48**: 206–213.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretsky I, Jaitin DA, David E, Keren-Shaul H, Mildner A, Winter D, Jung S, et al. 2014. Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345**: 943–949.

Lee BH, Tothova Z, Levine RL, Anderson K, Buza-Vidas N, Cullen DE, McDowell EP, Adelsperger J, Fröhling S, Huntly BJP, et al. 2007. FLT3 mutations confer enhanced proliferation and survival properties to

multipotent progenitors in a murine model of chronic myelomonocytic leukemia. *Cancer Cell* **12**: 367–380.

Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. 2016. An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nat Struct Mol Biol* **23**: 349–357.

Machanick P, Bailey TL. 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**: 1696–1697.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**: 495–501.

Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, Figueroa ME, Vasanthakumar A, Patel J, Zhao X, et al. 2011. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**: 11–24.

Quivoron C, Couronné L, Valle Della V, Lopez CK, 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.

Rasmussen KD, Jia G, Johansen JV, Pedersen MT, Rapin N, Bagger FO, Porse BT, Bernard OA, Christensen J, Helin K. 2015. Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev* **29**: 910–922.

Vassiliou GS, Cooper JL, Rad R, Li J, Rice S, Uren A, Rad L, Ellis P, Andrews R, Banerjee R, et al. 2011. Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice. *Nat Genet* **43**: 470–475.

Velasco-Hernandez T, Säwén P, Bryder D, Cammenga J. 2016. Potential Pitfalls of the Mx1-Cre System: Implications for Experimental Modeling of Normal and Malignant Hematopoiesis. *Stem Cell Reports* **7**: 11–18.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.