# Supplemental Information

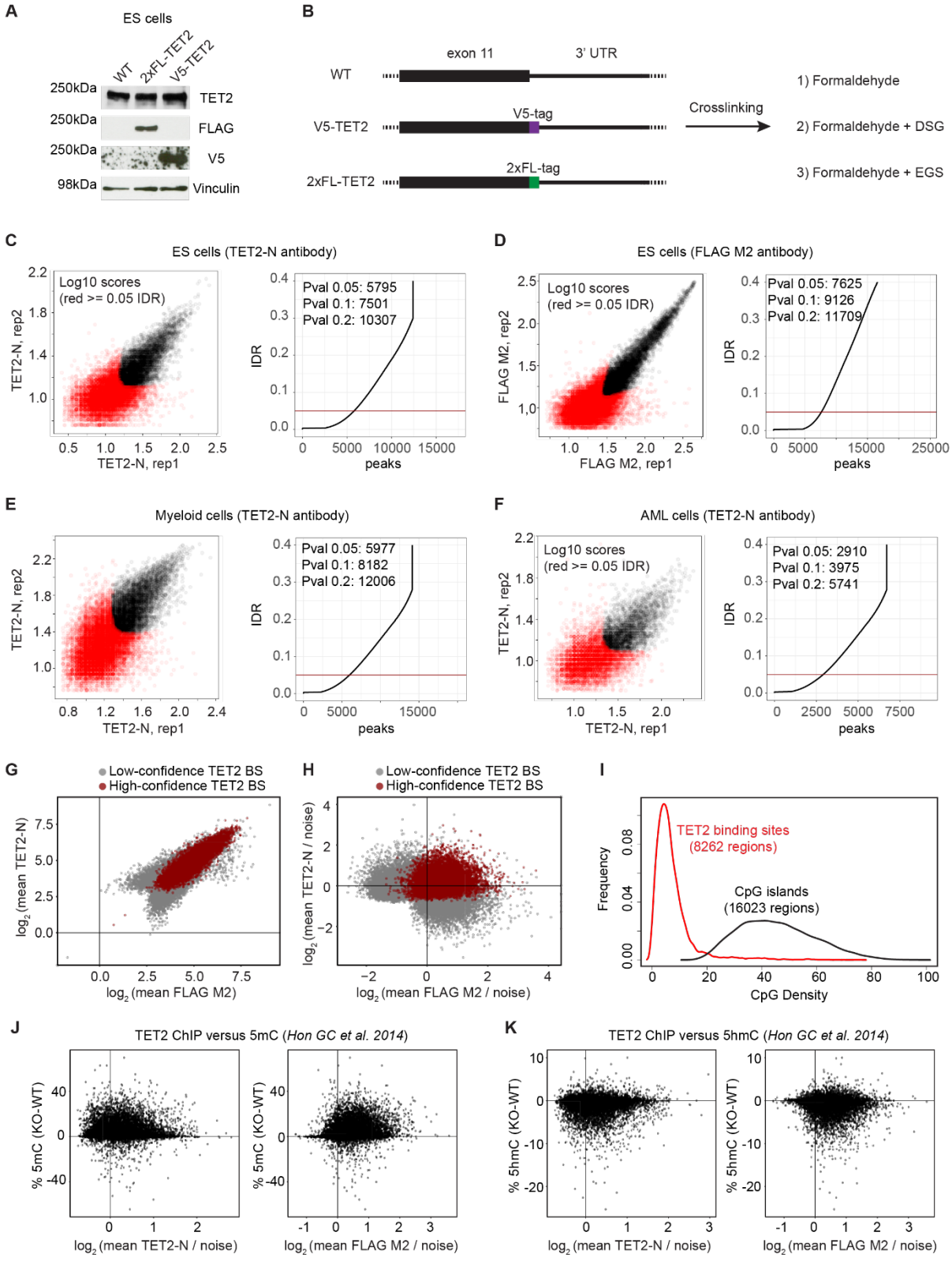
# TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells

**Kasper D. Rasmussen, Ivan Berest, Sandra Keβler, Koutarou Nishimura,**

**Lucía Simón-Carrasco, George S. Vassiliou, Marianne T. Pedersen,**

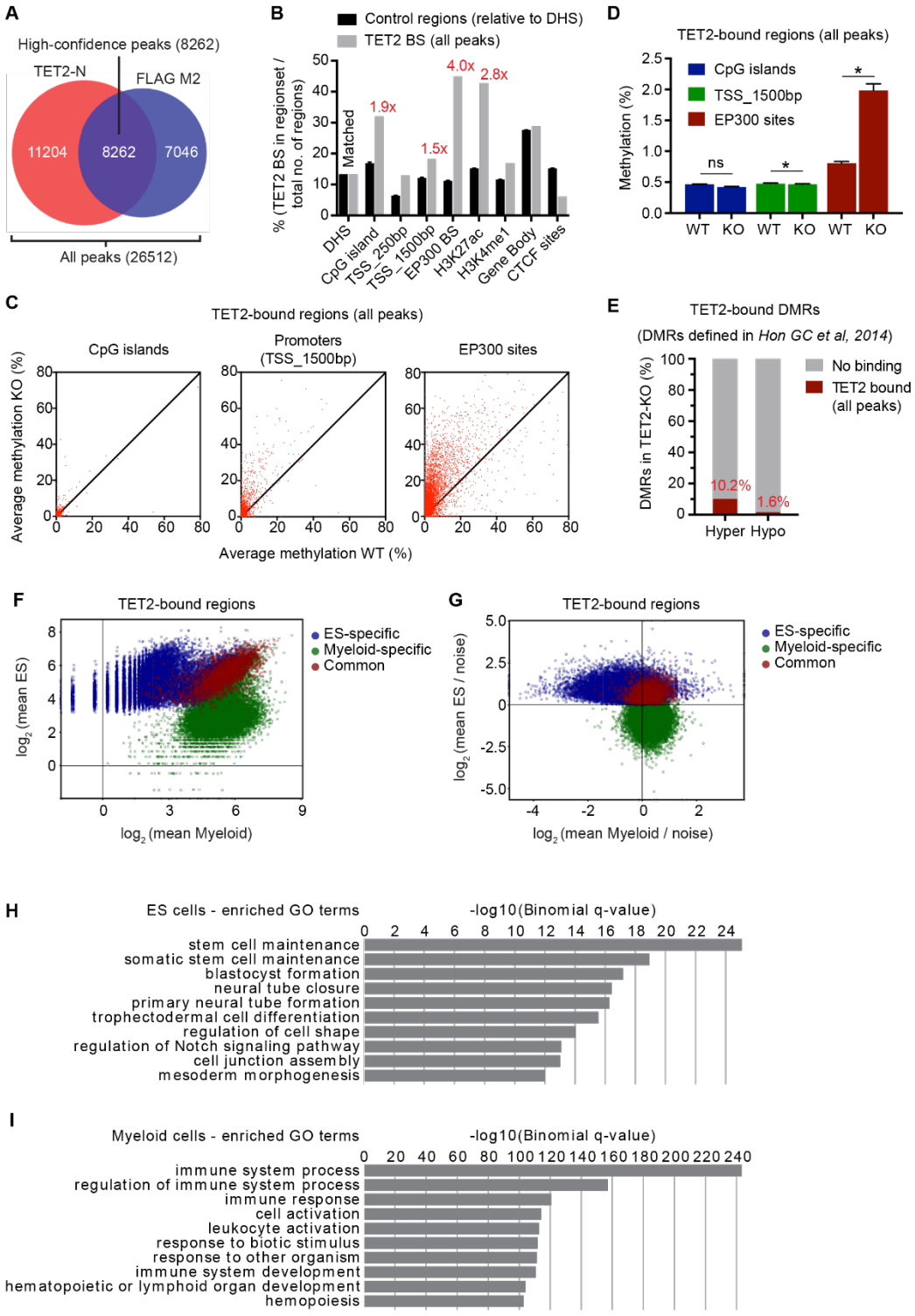
**Jesper Christensen, Judith B. Zaugg and Kristian Helin.**

**Table of content**




**Supplemental figure S1: TET2 chromatin immunoprecipitation and sequencing of wildtype and epitope-tagged cell lines. Related to Figure 1 and 2.**

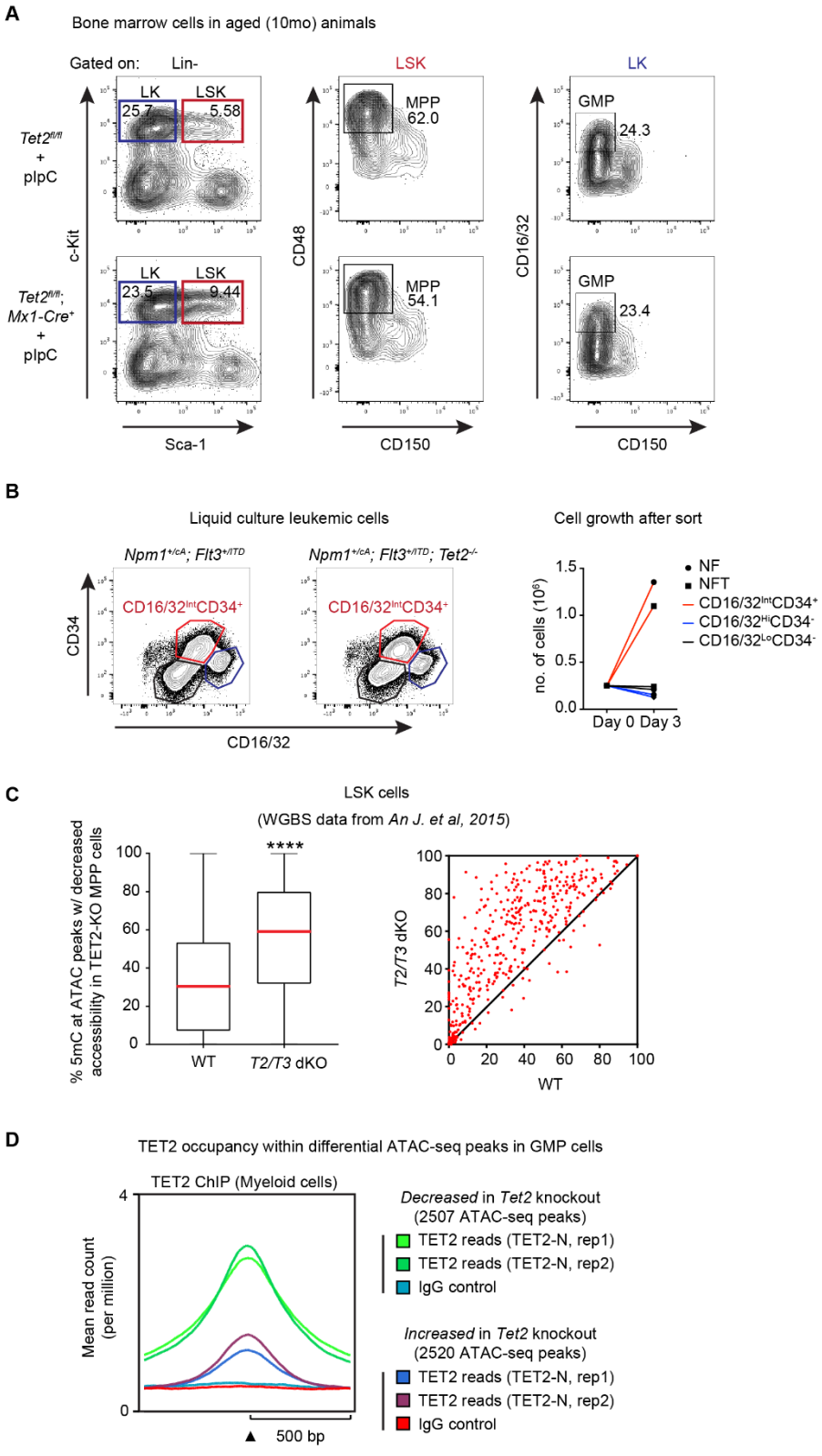
**(A)** Western blot of endogenously tagged TET2 in ES cells. **(B)** Schematic showing cell lines expressing endogenously tagged TET2 and crosslinking conditions used for ChIP-seq. **(C-D)** Irreproducibility Discovery Rate (IDR) analysis (Li et al. 2011) of replicate TET2 ChIP samples from ES cells generated with (C) TET2-N, or FLAG M2 (D) antibodies. **(E)** IDR analysis of replicate TET2 ChIP samples from myeloid hematopoietic cells. **(F)** IDR analysis of replicate TET2 ChIP samples from AML cells. The number of reproducible peaks at different confidence levels are shown. **(G)** Scatter dot plot showing correlation of $\log_2$ transformed normalized average TET2 ChIP-seq read counts in peaks in ES cells from TET2-N (y-axis) and FLAG M2 (x-axis) ChIP experiments. High-confidence (red dots) and low-confidence (grey dots) TET2 binding sites (BS) are marked. **(H)** Same as (G), but for $\log_2$ transformed fold changes of TET2 ChIP-seq reads over the background signal (noise) measured in *Tet2* knockout cells (for TET2-N) or empty cells (for FLAG M2). Variation in background noise between different replicates and antibodies increases the spread of the population of high-confidence peaks. **(I)** Plot showing density of CpG sites in high-confidence TET2 binding sites as well as CpG islands in the mouse genome. As expected from the TET2 genomic binding pattern (mostly found in promoter-distal enhancer regions) the CpG density is significantly lower as compared to CpG islands. The plot was generated using the cpgDensityPlot function in the RepiTools R package. **(J)** Scatter dot plot showing correlation of $\log_2$ transformed noise-subtracted TET2 ChIP-seq signal in TET2-N (left panel) or FLAG M2 (right panel) ChIP-seq experiments versus changes in average DNA methylation (5mC) in TET2 high-confidence binding sites and flanking regions (+/-250bp) (whole-genome bisulfite sequencing data (WGBS) from (Hon et al. 2014)). Only CpG sites covered by >10 reads were included in the analysis. **(K)** Same as (J), but for changes in average hydroxymethylation (5hmC) (whole-genome Tet-assisted bisulfite sequencing from (Hon et al. 2014)).




**Supplemental figure S2: Analysis of TET2 binding sites in ES and hematopoietic cells. Related to Figure 2 and 3.**

**(A)** Venn diagram as in Fig. 2A, but showing the origin of the low-stringency *all* peaks dataset in ES cells. **(B)** Histogram showing overlap of genomic regions as in Fig. 2E, but for *all* TET2 peaks. Within this peak set, TET2 binding is enriched in EP300 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 EP300 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 EP300 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 et al. 2014). **(F)** Scatter dot plot showing correlation of $\log_2$ average normalized TET2 ChIP-seq read counts in peaks in TET2-N ChIP experiments in ES cells (x-axis) and myeloid hematopoietic cells (y-axis). ES-cell specific (blue), Myeloid specific (green) and common (red) TET2-bound regions are marked. **(G)** Same as (F), but for $\log_2$ fold changes of TET2 ChIP-seq reads over the background signal (noise) measured in *Tet2* knockout in ES cells or myeloid hematopoietic cells. **(H)** Histogram showing enriched Gene Ontology (GO) terms identified by GREAT (McLean et al. 2010) analysis using promoter-distal (-1.5kb/+500bp from TSS) TET2 binding sites (TET2-N only) from ES cells. **(I)** Same as (H), but for myeloid hematopoietic cells.




**Supplemental figure S3. Native and malignant hematopoiesis with loss of TET2. Related to Figure 4.**

**(A)** Representative FACS plots showing the Lineage-negative, Sca-1-positive, cKit-positive (LSK) population in the bone marrow of aged animals (10 months old) 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)**

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. **(C)** Box (left) and scatter plot (right) showing average 5mC changes at ATAC-seq peaks with decreased chromatin accessibility in *Tet2* knockout MPP cells in wildtype and *Tet2*/*Tet3* double knockout LSK cells (An et al. 2015). Box plot shows median (red line) and 25$^{th}$ and 75$^{th}$ quantile, while scatter plot shows 5mC changes for individual ATAC-seq peaks. Only CpG sites covered by >10 reads were included in the analysis. ****, P < 0.0001 (Paired two-tailed Student`s t-test). **(D)** Plot showing the average TET2 ChIP-seq signal (measured in myeloid hematopoietic cells, described in (Rasmussen et al. 2015)) in regions that are associated with decreased or increased accessibility in *Tet2* knockout GMP cells. The TET2 ChIP signals in both replicate experiments are significantly higher in ATAC-seq peaks with reduced accessibility compared to ATAC-seq peaks that gain accessibility in *Tet2* knockout GMP cells.

**Supplemental Experimental Procedures**

**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 analyzed using FlowJo software (Tree Star inc.).

**Western blotting and antibodies**

Total cells lysates were made by boiling in 2x LSB sample buffer and Western blotting was carried out according to standard protocols with a primary rabbit antibody to TET2 (TET2-N, Helin lab), Vinculin (Sigma, F7425), V5-epitope (Abcam, ab15828), and FLAG-epitope (Sigma, F7425). The polyclonal antibody to TET2 (TET2-N) was generated by immunizing rabbits with affinity-purified bacterially expressed murine GST–TET2N (amino acids 1–300). The antibodies were absorbed on GST-coupled cyanogen bromide-activated Sepharose (GE Healthcare) and subsequently affinity purified using Sepharose coupled with GST–TET2N.

Bands on nitrocellulose membranes were visualized using Super Signal West Pico chemiluminescent ECL substrate (ThermoFisher Scientific) by exposing on Amersham hyperfilm ECL films (GE healthcare).

**RNA-seq library generation and analysis**

A proportion of total RNA (2 ng) isolated using RNeasy Micro Kit (Qiagen) 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 analyzed using DESeq2. Four biological replicates were used for the DESeq2 analysis. Gene counts were normalized by variance stabilizing transformation (VST). 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).

**ChIP-seq library generation**

ES cells or *in vitro* cultured 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) and incubated at a rotating wheel for 30 min to allow equilibration to room temperature. Then, formaldehyde (Sigma) 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 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, chromatin was 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) (as described above) was incubated with 300µg chromatin (measured by Bradford assay) overnight. The chromatin-antibody complexes were captured in a 3h incubation with protein-G Sepharose beads (GE healthcare). For immunoprecipitation of 2xFL-TET2, 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 de-crosslinked 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 (DYKDDDDK, Peptide 2.0). The eluted fractions were pooled and de-crosslinked 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 19,466 and 15,308 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 (Zhang et al. 2008). Peaks were called with default parameter settings and a 0.05 q-value cutoff 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 (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 with default settings.

**ATAC-seq library generation and preprocessing**

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 approximately 25 million reads per sample.

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 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. 2018). 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).

Briefly, we first scanned the genome for TF binding sites using the mouse PWMs deposited in the database HOCOMOCO v10 (without two retracted motif for CDX4 and EVX1 and combined with 6 PWMs for CDX4 and EVX1 from the recent methyl-SELEX study (Yin et al. 2017)) using PWMScan (Ambrosini et al. 2018) to identify *in silico* predicted binding sites for each TF. For each TF, we then extracted the signal around binding sites (+/- 100bp around the core motif) and calculated a fold-change of ATAC-seq reads 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%). To calculate the mean difference, we compared, for each bin, the distribution of fold-changes for binding sites of one TF to the background distribution of fold-changes across all other TFs. 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. In order to estimate variance of the resulting z-scores we employ bootstrap approach (n=1000) with resampling of the bin-specific data, followed by incorporation of the possible co-variances between bins (see more details in (Berest et al. 2018)). We centralized the distribution of the resulted z-score values using locfdr package in R and calculated p-values out of z-scores based on the TF-specific variance calculated above. Finally, we adjusted the p-values using Benjamini-Hochberg multiple testing correction.

**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 6A). 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 considered. Only expressed TFs that had more than 50 target genes and an FDR < 0.05 in the *diffTF* analysis were considered for the correlation.

**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. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw343.

Ambrosini G, Groux R, Bucher P. 2018. PWMScan: a fast tool for scanning entire genomes with a position-specific weight matrix ed. J. Hancock. *Bioinformatics* **34**: 2483–2484. https://academic.oup.com/bioinformatics/article/34/14/2483/4921176 (Accessed November 26, 2018).

An J, González-Avalos E, Chawla A, Jeong M, López-Moyado IF, Li W, Goodell MA, Chavez L, Ko M, Rao A. 2015. Acute loss of TET function results in aggressive myeloid cancer in mice. *Nat Commun* **6**: 10071. http://www.nature.com/articles/ncomms10071.

Berest I, Arnold C, Reyes-Palomares A, Palla G, Rasmussen KD, Helin K, Zaugg J. 2018. Quantification of differential transcription factor activity and multiomic-based classification into activators and repressors: diffTF. *bioRxiv* 368498. https://www.biorxiv.org/content/early/2018/07/13/368498.full.pdf+html (Accessed July 16, 2018).

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. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=24097267&retmode=ref&cmd=prlinks.

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. http://linkinghub.elsevier.com/retrieve/pii/S1097276514006819.

Kumasaka N, Knights AJ, Gaffney DJ. 2016. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat Genet* **48**: 206–213. http://www.nature.com/articles/ng.3467.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359. http://www.nature.com/articles/nmeth.1923.

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. http://science.sciencemag.org/content/345/6199/943.abstract.

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. http://www.nature.com/articles/nsmb.3180.

Li Q, Brown JB, Huang H, Bickel PJ. 2011. Measuring reproducibility of high-throughput experiments. *Ann Appl Stat* **5**: 1752–1779. http://projecteuclid.org/euclid.aoas/1318514284 (Accessed November 22, 2018).

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. http://www.nature.com/articles/nbt.1630.

Rasmussen KD, Jia G, Johansen J V, 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. http://genesdev.cshlp.org/lookup/doi/10.1101/gad.260174.115.

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, et al. 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**: eaaj2239. http://www.sciencemag.org/lookup/doi/10.1126/science.aaj2239.

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. http://genomebiology.biomedcentral.com/articles/10.1186/gb-2008-9-9-r137.