Online Methods

Patient samples

Genomic DNA was extracted from bone marrow aspirates and occasionally peripheral blood from healthy donors and patients with MDS, MDS/MPN, primary and secondary AMLs. Samples were collected at the Cleveland Clinic according to protocols approved by Institutional Review Boards (IRBs). Clinical features and other detailed information pertaining to the patient samples are summarized in **Suppl. Table 1**. Mutants are defined as follows: Homozygous, duplication of mutation due to somatic CN-LOH affecting 4q24; hemizygous, mutant *TET2* on one allele and *TET2* deletion on the other allele; heterozygous, mutant *TET2* on one allele, apparent wild type *TET2* on the other allele; biallelic heterozygous, different mutation on each allele. One sample from a healthy donor was excluded in the dot blot analysis of **Fig. 2c** and **Suppl. Fig. 7b** due to the presence of a white precipitate that blocked bisulfite conversion.

SNP-A analysis

DNA was extracted from patient marrow specimens and when sample material allowed, $CD3^+$ cell fractions were isolated using RoboSep (StemCell Technologies, Vancouver, Canada). The Affymetrix GeneChip Human Mapping 250K Array and Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA) were used for SNP-A analysis as previously described[6]. Signal intensity was analyzed and SNP calls determined using GeneChip Genotyping Analysis Software Version 4.0 (GTYPE). Copy number variations (CNV) and areas of CN-LOH were investigated using a Hidden Markov Model and CN Analyzer for Affymetrix GeneChip Mapping 250K arrays (CNAG v.3.0). Genotyping console v2.0 (Affymetrix) was used for analysis of 6.0 arrays. We excluded germ-line encoded copy number variations (CNV) and non-clonal areas of UPD from further analysis by utilizing a bioanalytic algorithm based on lesions identified by SNP-A in an internal control series (N=1003) and reported in the Database of Genomic Variants (http://projects.tcag.ca/variation/). Whenever possible paired germ line DNA derived from $CD3^+$ cells was used.

TET2 mutational screening

Screening for mutations in *TET2* was carried out using direct genomic sequencing as previously described[6]. Briefly, PCR primers were designed to amplify and sequence *TET2* coding region (exons 3-11). When needed, multiple set of primers overlapping by 100 bp were used to ensure complete coverage. Genomic DNA from bone marrow (occasionally from peripheral blood) was used for PCR amplification followed by purification using Montage®Cleanup Kit (Millipore, Billerica, MA) and sequencing using ABI 3730xl DNA analyzer. All *TET2* mutations were detected by bidirectional

sequencing and scored as pathogenic mutations on the basis of the observation that they were not detected in normal samples or non-clonal $CD3^+$ cells and/or were reported by other groups as mutants. All mutations were annotated according to the coding sequence using *TET2* isoform A (NM_001127208). Mutations in all intron-exon boundaries (acceptor/donor sites) were included, all other intron mutations were not scored. Homozygous mutations include patients with CN-LOH spanning the 4q24 region, whereas hemizygous configurations have deletions of the 4q24 region.

Tet2 expression and knockdown plasmids

The open reading frame of mouse *Tet2* (see **Suppl. Figs. 1, 2**) was amplified using cDNA from V6.5 mouse embryonic stem cells; its sequence differs from that in the public databases and is provided in **Suppl. Fig. 1**. The Myc epitope was inserted at the N-terminus by PCR and then Myc-*Tet2* was inserted into the EcoRI site of **pEF1a** (Invitrogen). Site-directed mutagenesis of Myc-*Tet2* was performed using QuikChange Mutagenesis kit (Stratagene). The sequences were confirmed by automated DNA sequencing. The EcoRI fragment of wild type and mutant Myc-*Tet2* was cloned into the EcoRI site of **pEF1a-IRES-CD25** vector. RNA interference (RNAi)-mediated knockdown of *Tet2* was achieved by **MSCV-LTRmiR30-PIG** (LMP, OpenBiosystems) retroviral vectors. Sequences of the shRNA hairpins were as follows:
Tet2 shRNA: 5'-
TGCTGTTGACAGTGAGCGCCACTACTAACTCCACCCTAAATAGTGAAGCCACAGATGTATTTAGG
GTGGAGTTAGTAGTGATGCCTACTGCCTCGGA-3',
Scramble: 5'-
TGCTGTTGACAGTGAGCGCGCACGAAGCGTATGGATACAATAGTGAAGCCACAGATGTATTGTA
TCCATACGCTTCGTGCTTGCCTACTGCCTCGGA-3'.

Immunocytochemistry

HEK293T cells (~4 x 10^5) plated on sterile coverslips in 24-well plates were cultured overnight before transient transfection with wild type or mutant **Myc-*Tet2*-pEF1** plasmids or empty vector using Lipofectamine 2000 (Invitrogen) as per manufacturer's instructions. After 48 hr, cells were fixed with 4% paraformaldehyde in PBS for 15 min and permeabilized with 0.2% Triton X-100 in PBS for 15 min at room temperature (RT). Subsequently, DNAs were denatured with 2N HCl at RT for 30 min and neutralized with 100 mM Tris-HCl buffer (pH 8.0) for 10 min. After washing with PBS, cells were blocked with 1% BSA, 0.05% Tween-20 in PBS at RT for 1 hr. Then, rabbit anti-5-hmC polyclonal antibody (diluted at 1:1,000) for 5-hmC staining or mouse anti-5-mC antibody (clone 162 33 D3; diluted at 1:2,500 dilutions, Calbiochem, San Diego, CA) and rabbit anti-Myc antibody (diluted at 1:1,000 dilutions, Sigma) for 5-mC staining were added into the blocking buffer for 2 hr at RT. After extensive washing with PBS, Cy2-conjugated anti-rabbit IgG (for 5-hmC staining, Jackson Immunoresearch) or Cy2-conjugated anti-rabbit IgG and Cy3-conjugated anti-mouse IgG (for 5-mC

staining, Jackson Immunoresearch) were added in blocking buffer for 1 hr in the dark. For detection of Myc epitope in 5-hmC staining, cells were additionally incubated with anti-Myc antibody (9E10) conjugated with Alexa Fluor 555 (Millipore) in blocking buffer for 1 hr at RT. DNA was stained with 250 ng/ml of 4',6-diamidino-2-phenylindole (DAPI) and mounted in SlowFade Gold antifade reagent (Molecular Probes, Eugene, OR). Images were obtained with Zeiss Axiovert 200 microscope using OpenLab imaging software (Improvision, Coventry, UK).

Transfection and sorting of HEK293T cells expressing hCD25

HEK293T cells were transfected with wild type and mutant **Myc-*Tet2*-IRES-hCD25-pEF1a** vectors using calcium phosphate co-precipitation method. Control cells were transfected with empty vector expressing hCD25 alone. After 48 h, CD25+ cells were purified by magnetic isolation. Cells were resuspended at 5 x 10^7-10^8 cells/ml and incubated with anti-hCD25-PE antibody (BD Biosciences, 1:200) in 1X MACS buffer (1X PBS, 2% FBS, 2 mM EDTA) for 20 min at 4ºC in the dark. After washing with 1X MACS buffer, cells were then stained with anti-PE microbeads (Miltenyi Biotech) for 20 min at 4ºC in the dark. Then, cells were washed twice with 1X MACS buffer and resuspended in 1 ml ice-cold 1X MACS buffer. $CD25^+$ cells were sorted using LS column (Miltenyi Biotech). Input, flow-through and column eluates were analyzed by FACS to ensure the enrichment of $CD25^+$ cells in eluates.

Quantitative analysis of 5-hmC and CMS levels using dot-blot

Cells were incubated with 200 µg/ml proteinase K (Roche) overnight at 55^{o}C and genomic DNA was isolated with equal volumes of phenol, phenol: chloroform: isoamyl alcohol (25:24:1) and chloroform: isoamyl alcohol (24:1) and then precipitation with 2 volumes of ethanol. For CMS detection, genomic DNA from cell lines, bone marrow or blood (if it contained >20% immature myeloid cells) was treated with sodium bisulfite using the EpiTect Bisulfite kit (QIAGEN). DNA samples were denatured in 0.4 M NaOH, 10 mM EDTA at 95ºC for 10 min, and then neutralized by adding an equal volume of cold 2 M ammonium acetate (pH 7.0). 2-fold serial dilutions of the denatured DNA samples were spotted on a nitrocellulose membrane in an assembled Bio-Dot apparatus (Bio-Rad) according to manufacturer's instructions. The membrane was washed with 2X SSC buffer, air-dried and vacuum-baked at 80°C for 2 hrs, then blocked with 5% non-fat milk for 1 hr and incubated with anti-5-hmC or anti-CMS antibody (1:1,000) overnight at 4^{o}C. After incubating with HRP-conjugated anti-rabbit IgG secondary antibody, the membrane was visualized by enhanced chemiluminescence. To ensure equal spotting of total DNA on the membrane, the same blot was stained with 0.02% methylene blue in 0.3 M sodium acetate (pH 5.2). In **Figs. 4a, b**, which incorporate data from **Fig. 2** and **Suppl. Fig. 6**, we compared results obtained in the different experiments by using the normalisation procedure described in the section *Statistical Analysis* below.

Western blotting

Cells were lysed with RIPA buffer (150 mM NaCl, 50 mM Tris-HCl, pH 8.0, 1% Triton X-100, 0.5% sodium deoxycholate and 0.1% SDS) supplemented with protease inhibitor cocktail (Roche) and incubated on ice for 30 min. Cell debris was removed by centrifuging at 12,000 rpm for 15 min at 4 °C. The protein concentration was measured by Bradford protein assay. Samples were mixed with SDS sample buffer and boiled for 5 min. Whole cell lysates were separated by 7.5% SDS-PAGE and transferred onto PVDF membranes. Proteins were detected by immunoblotting in TBST (150 mM NaCl, 10 mM Tris-Cl, pH 8.0, 0.5% Tween-20) containing 5% low-fat milk and anti-Myc (9E10) or anti-Actin (Sigma) antibodies, followed by incubation with HRP-conjugated secondary antibody.

RNA purification, reverse transcription and quantitative RT-PCR

Total RNA was prepared with TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Equivalent amounts (1-1.5 μg) of total RNA were reverse-transcribed using SuperScript III reverse transcriptase (Invitrogen). Diluted cDNAs were analyzed by real-time PCR using StepONE plus real-time PCR system (Applied Biosystems) and FastStart Universal SYBR Green Master kit (Roche). The program was set as 95°C for 10 min, 40 cycles of 95°C for 15 sec, 60°C for 1 min, and a melt curve analysis step at the end. Data were analyzed by StepONE plus real-time PCR software. The level of gene expression was normalized to *Gapdh*. Primer sequences are as follows: *Tet1* forward 5'-GAGCCTGTTCCTCGATGTGG-3', *Tet1* reverse 5'-CAAACCCACCTGAGGCTGTT-3', *Tet2* forward 5'-AACCTGGCTACTGTCATTGCTCCA-3', *Tet2* reverse 5'-ATGTTCTGCTGGTCTCTGTGGGAA-3', *Tet3* forward 5'-TCCGGATTGAGAAGGTCATC-3', *Tet3* reverse 5'-CCAGGCCAGGATCAAGATAA-3', *Gapdh* forward 5'-GTGTTCCTACCCCCAATGTGT-3' and *Gapdh* reverse 5'-ATTGTCATACCAGGAAATGAGCTT-3'.

ES cell culture

V6.5 mouse ES cells were maintained in culture as previously described [1]. Briefly, V6.5 mouse ES cells were cultured on primary mouse embryonic fibroblasts that treated with mitomycin C. The culture medium contains DMEM knockout (Invitrogen, Carlsbad, CA), 15% ES FBS (Omega Scientific, Tarzana, CA), 0.1 mM non-essential amino acids (invitrogen), 2 mM L-glutamine (Invitrogen), 0.1 mM β-mercaptoethanol (Invitrogen), 50 units/ml penicillin/streptomycin (Invitrogen) and 1,000 U/ml LIF (Chemicon). Total RNA purifiation, cDNA synthesis and qRT-PCR was performed as described above.

Flow cytometry and cell sorting

Bone marrow cells were flushed out of femurs and tibias of mice, and red blood cells were depleted by incubation in with ACK lysis buffer (0.15 M NH_4Cl, 10 mM $KHCO_3$ and 0.1 mM EDTA) at RT for 1

min. Thymocytes were prepared by mincing thymuses onto a 70 μm cell strainer (BD Biosciences). Cells were stained with monoclonal antibodies in various combinations in PBS containing 1% heat-inactivated FBS and 0.1% (w/v) sodium azide. Biotinylated antibodies were revealed by PerCP-Cy5.5- or APC-conjugated streptavidin (BD Biosciences). Flow cytometric analyses and cell sorting were performed using FACSAria II flow cytometer (BD Biosciences) and data were analyzed using FlowJo software. The scheme for hematopoietic sorting is described in **Suppl. Fig. 8**.

Lin$^-$ cell purification, retrovirus transduction and *in vitro* differentiation

For enrichment of Lin$^-$ cells, bone marrow cells prior to red blood cell depletion were incubated with Biotin-conjugated Mouse Hematopoietic Lineage Panel (eBioscience) at 4^{o}C for 20 min. After washing with 1X MACS buffer, cells were incubated with anti-Biotin microbeads (Miltenyi Biotech) at 4°C for 20 min and Lin$^-$ cells were purified by magnetic separation using LS columns. To produce retroviruses, HEK293T cells were co-transfected with retroviral constructs and packaging construct, Ecopac (Cell Genesys) by using calcium phosphate co-precipitation method. After 48 hr, viral supernatant was collected and filtered through 0.45 μm filters before freezing at -80^{o}C. Upon purification, Lin$^-$ BM cells were cultured overnight in complete RPMI (20% FBS, 100 U/ml penicillin, 100 U/ml streptomycin, 50 μg/ml β-mercaptoethanol, 1% Glutamax (GIBCO), 1% non-essential amino acid (GIBCO) and 1% sodium pyruvate (GIBCO)) supplemented with 6 ng/ml IL-3, 10 ng/ml IL-6, 100 ng/ml SCF and 50 ng/ml Flk2/Flt3 ligand (all from R&D Systems). For transduction, equal volume of viral supernatant were added to 6-well plates containing cells, followed by spinoculation in the presence of cytokines and 8 μg/ml polybrene at 2,500 rpm for 90 min. After 36-48 hr, cells were washed and further cultured in media containing 50 ng/ml SCF, 2 μg/ml puromycin and combinations of GM-CSF, M-CSF and G-CSF (all from R&D Systems). After 4 days, cell were harvested and analyzed for GFP expression and myeloid differentiation by flow cytometry.

Statistical analysis

To investigate the impact of *TET2* mutations on 5-hmC level, we measured 5-hmC levels in triplicate for samples assigned to one of three different groups: healthy donors, mutant *TET2* and wild type *TET2*. For **Figure 2c**, triplicate measurements, computed from the linear portion of the standard curve for each sample (**Suppl. Fig. 7b**) were summarized to one value using the median as a robust estimator. Based on Kolmogorov-Smirnov tests for normality, no group was normally distributed (*p*-values < 10^{-3}). Hence, we employed two-sided Wilcoxon Rank Sum tests to compare the different groups.

For **Figure 4a, b**, normalised 5-hmC values were plotted. For each of the five different experiments shown in **Fig. 2** and **Suppl. Fig. 6**, we rescaled all 5-hmC values by subtracting the lowest value observed in that experiment from all the other values, then normalised the data by dividing the resulting values by the highest adjusted 5-hmC value. That is, for each experiment, we set the lowest

5-hmC value to 0 and the difference (highest minus lowest) to 1, and scaled all intermediate experimental values accordingly. The obtained values range from 0 to 1 and are comparable under the assumption that each experiment only differed by an offset and a multiplicative factor, which can be estimated from the data.

Methylation analysis

Methylation detection was performed on bisulfite-treated DNA using the Illumina® Infinium® 27k array (Illumina, San Diego, CA) according to the manufacturer's protocol and as previously described[26]. Bisulfite conversion of 1 μg DNA of patients' samples was performed using EZ DNA methylation kit (Zymo Research, Orange, CA). The Infinium consists of paired probes for each of 27,578 CpG dinucleotides in the genome, corresponding to methylated and unmethylated sites. Through allele-specific ligation and extension, PCR templates were generated and amplified using fluorescently-labeled universal primers. Resulting PCR products were hybridized to a bead array bearing complementary sequences. These hybridization targets contained a fluorescent label denoting methylated or unmethylated status for a given locus. Methylation status was calculated from the ratio of the methylation-specific and demethylation-specific fluorophores (β-value) using BeadStudio® Methylation Module (Illumina, San Diego, CA). Results were normalized to the background fluorescence for each array. Several different control oligonucleotides were used to ensure the highest quality data including allele-specific extension control, bisulfite conversion controls, extension gap controls, gender controls, first hybridization controls, second hybridization controls, negative controls and contamination detection controls.

Preliminary analyses showed that methylation sites on Y and X chromosomes show inconsistent methylation status with respect to gender. As this is a known issue (communication with Illumina), we decided to remove those sites from the analysis. All calculations are based on the β values, which correspond to the methylation status of a site ranging from 0 to 1, returned by Illumina's BeadStudio software. We test sites for differential methylation using an empirical Bayes approach employing a modified t-test (limma). The false discovery rate (FDR) is controlled at a level of 5% by the Benjamini-Hochberg correction.