Supplementary Figures and Table

Supplementary text

Additional analysis of patients with wildtype or mutant TET2. Essentially all patients with *TET2* mutations exhibited low genomic levels of 5-hmC, irrespective of the stage of the disease (low risk versus advanced, gauged by percentage of blasts in the bone marrow or risk assessment by International Prognostic Scoring System); this most likely reflects the fact that in bone marrow of these disorders (MDS and in particular CMML), the majority of cells are derived from malignant clones.

Low 5-hmC levels were seen in a considerable fraction of patients with wild type *TET2* (**Figs. 2c**, **4a, 4b**). It is unlikely that these patients have loss-of-function of TET1 or TET3: both *TET1* and *TET3* loci have been sequenced extensively in several laboratories, including ours, and no mutations or lesions identified by SNP-A karyotyping were found [3,7] (**Suppl. Table S1**, and data not shown). Moreover, the chromosomal region containing *TET1* (10q21) is only rarely subject to deletions or loss of heterozygosity in myeloid malignancies, unlike the 4q24 region that harbours *TET2*.

To test the possibility that the *TET2* alleles of some patients might harbor undiscovered mutations in *cis*-regulatory regions that diminish *TET2* mRNA expression, we performed quantitative RT-PCR on mRNA extracted from healthy controls and patients with mutant *TET2*, patients with wild type *TET2* but low 5-hmC, and patients with wild type *TET2* and high 5-hmC (n=6 for each group). This analysis did not reveal any systematic differences in *TET2* expression in the four groups (A.M.J., Y. H., J. P. M. and A. R., data not shown).

The most interesting possibility is that patients with wild type *TET2* but low 5-hmC bear mutations in other proteins involved in TET2-mediated catalysis. These could include partner proteins that associate with TET2, metabolic enzymes implicated in the biogenesis of TET2 co-factors, or regulatory proteins controlling the appropriate targeting of TET2 to relevant genomic regions / TET2 target genes.

A final possibility is that CpG hypomethylation is the primary event in a subset of tumours with wild type *TET2* but low 5-hmC. By diminishing the levels of the TET2 substrate, 5-mC, this might cause a secondary decrease in the levels of 5-hmC.

Expression pattern of Tet family members during haematopoietic development. During erythroid development, *Tet2* expression was progressively downregulated as LSK differentiated into megakaryocyte-erythroid progenitors (MEPs) and then into early erythoblasts ($Ter\text{-}119^+$) (**Suppl. Fig. 9a,** *middle panel***)**. *Tet2* mRNA was also strongly expressed in common lymphoid progenitors (CLPs), gradually increased in early B cell precursors in the bone marrow, then declined slightly in mature B cells; in contrast, it was progressively downregulated during early T cell development in the thymus

but strongly induced at the double-positive (DP) and single-positive (SP) stage (**Suppl. Fig. 9a**). Tet1 showed a different expression pattern: *Tet1* mRNA levels progressively declined during myeloid differentiation as well as during the later stages of B cell differentiation, but were high in DP and SP thymocytes, the terminal stages of T cell thymic differentiation (**Suppl. Fig. 9a,** *top panel***)**. Another member of the Tet family, Tet3, showed a progressive decline in mRNA expression during myeloid differentiation. *Tet3* mRNA was also expressed in CLPs and gradually increased during early B cell development, and then declined in mature B cells (**Suppl. Fig. 9a,** *bottom panel*).

In parallel, we analysed Tet mRNA expression in mouse embryonic stem (ES) cells. Undifferentiated ES cells expressed high levels of *Tet1* mRNA, moderate levels of *Tet2* mRNA, and no *Tet3* mRNA. After exposure for 4 days to 1 μM retinoic acid (RA), *Tet1* and *Tet2* mRNA levels declined to very low levels whereas *Tet3* mRNA levels increased by several-fold.

We also assessed 5-hmC levels by anti-CMS dot blot after bisulfite treatment of DNA. 5-hmC levels were high in LSK progenitors, mature Gr-1$^+$Mac-1$^+$ granulocytes and Gr-1$^-$Mac-1$^+$ monocyte populations, but low in Ter-119$^+$ erythroid cells (**Suppl. Fig. 9b**). In B cell lineages, the 5-hmC level was high in pro-B cells but decreased in pre-B cells, followed by gradual increase at subsequent development stages (immature and mature B cells; **Suppl. Fig. 9b**). In the thymus, 5-hmC level was high in single-positive (SP) CD4$^+$ and CD8$^+$ thymocytes, but relatively lower at immature single positive (ISP) and double-positive (DP) stages (**Suppl. Fig. 9c**).

Suppl. Table 1. Summary of clinical features and *TET2* mutational status in patients and healthy control individuals who provided genomic DNA for measurement of 5-hmC levels in **Fig. 2** and **Suppl. Fig. 7**. Samples from *Expt. 1* were analysed in **Fig. 2a-c**, samples from *Expts. 2-4* in **Suppl. Fig. 7**, and samples from *Expt. 5* in Fig. **2d**.

Suppl. Table 2. Sites of differential methylation in comparisons of bone marrow/ peripheral blood samples from patients with low 5-hmC versus healthy controls**.**

Suppl. Figure. 1. Protein sequence of the mouse Tet2 C-terminal region encompassing its catalytic domain, compared with the protein sequences predicted from Ensembl (ENSMUSG00000040943) and GenBank (NM_001040400). The cysteine-rich (Cys-rich) region (amino acids 1,042-1,241 in our sequence) and double stranded beta-helix region (amino acids 1,242-1,919 in our sequence) are indicated. The region differing from that in the public databases is enclosed in a red rectangle. The amino acids targeted for mutagenesis are indicated. The entire nucleotide sequence of the Tet2 ORF cloned from cDNA of V6.5 mouse ES cells is available upon request.

Suppl. Figure 2. Protein sequence alignment of mouse Tet2 with human TET2. Leukemia-

associated amino acid substitutions in the human TET2 sequence, and the corresponding amino acids targeted for mutagenesis in the mouse Tet2 sequence, are indicated.

Suppl. Figure 3. The enzymatic activity of Tet2 is impaired by leukemia-associated mutations.

a, Schematic representation of additional missense mutations in *TET2*. Five leukemia-associated mutations discovered in patients and the corresponding substitutions introduced into murine *Tet2* are indicated.

b, Immunocytochemical detection of Myc-Tet2 (red) and 5-hmC (green) in HEK293T cells transiently transfected with wild type and mutant Tet2. Staining was performed as in **Fig. 1b**.

c, Detection of Myc-Tet2 (green) and 5-mC (red) in HEK293T cells (staining performed as in **Fig. 1c**).

d, Western blot analysis of purified ($CD25^+$) HEK293T cells expressing wild type and mutant Myc-Tet2. Tet2 was expressed and detected as described in **Fig. 1d**.

e, Dot-blot analysis of 5-hmC levels in HEK293T cells over-expressing wild type or mutant Tet2. CMS levels in bisulfite-treated DNA were quantified as described in **Fig. 1e**. As positive and negative controls, we included DNA from $CD25^+$ HEK293T cells transfected with TET1 catalytic domain (TET1-CD) or TET1-CD with mutations in the HxD motif (TET1-CD-HxDmut)[1].

Suppl. Figure 4. Quantification of fluorescence intensity in transfected cells. The fluorescence signal intensity of 5-hmC and 5-mC in HEK293T cells transfected with wild type or mutant *Tet2* in **Figs. 1b, c** (a, b) and **Suppl. Figs. 3b, 3c** (c, d) was measured using Photoshop CS4 software and plotted using GraphPad Prism 5. The horizontal bar indicates the mean for each group.

Suppl. Figure 5. Specificity and density dependence of polyclonal antibodies against 5-hmC or its adduct after bisulfite treatment, cytosine 5-methylsulfonate (CMS).

a, Sodium bisulfite converts 5-hmC into cytosine-5-methylenesulfonate (CMS)[14,15].

b, Specificity of the anti-5-hmC and anti-CMS antisera. Rabbits were immunized with 5-hmC or CMS conjugated to KLH. Antisera collected from these immunized rabbits were tested in the dot-blot assay for their ability to selectively detect distinct cytosine modifications on serially-diluted (0.05 pmol to 3 pmol) lambda (λ) phage DNA, lambda (λ) phage DNA methylated *in vitro* at CpG sites with SssI (λ^{Me}), or DNA from 5-hmC-containing T4* bacteriophage (produced by growth on the appropriate *E. coli* host[1]) before (T4*) and after bisulfite treatment (T4*-BiS). Both antisera were highly specific for the respective cytosine modifications.

c, Density dependence of anti-5-hmC and anti-CMS antisera. Oligos with different densities of 5-hmC were produced by PCR-amplifying a 201 bp double-stranded oligo containing a total of 81 cytosines using dCTP, 5hmdCTP, or mixtures of dCTP with 20%, 5%, and 1.25% 5hmdCTP. Two-fold dilutions

of the amplified DNA were assayed by dot blot as in **b**. The data are plotted as dot blot intensity normalized to the total amount of 5-hmC in the spot (i.e. 1 pmol of an oligo in which all 81 cytosines are replaced with 5-hmC contains 81 pmol 5-hmC, equivalent to 5 pmol of an oligo generated with a 20:80 mixture of 5-hmC:C, that contains ~16 5-hmC on average). *Top*, When the anti-5-hmC antiserum is used to detect 5-hmC directly, a 5-fold drop in 5-hmC density results in a >30-fold drop in detection efficiency. *Bottom*, If instead the oligos are treated with sodium bisulfite before performing the dot blot, an 80-fold change in the density of 5-hmC (CMS) produces only an ~10-fold drop in detection efficiency with the anti-CMS antiserum, and this is partly due to the fact that when PCR amplification is performed at low 5-hmC:C ratios, many of the resulting PCR fragments lack 5-hmC entirely. Thus unlike the anti-5-hmC antiserum which reacts efficiently only with oligos containing high densities of 5-hmC, the anti-CMS antiserum efficiently recognizes oligos with even a single 5-hmC.

Suppl. Figure 6. Analysis of 5-hmC and 5-mC levels in HEK293T cells expressing wild type or mutant Tet2.

a-b, Genomic DNA was extracted from purified $CD25^+$ HEK293T cells over-expressing wild type or mutant Tet2 (see **Fig. 1**) and blotted onto nitrocellulose membrane after two-fold serial dilution. 5-hmC was detected with antisera against 5-hmC (**a**, see **Suppl. Fig. 5b**, *left panel*) or 5-mC (**b**, Calbiochem, NA#81).

c-d, Dot-blot analysis to quantify levels of 5-hmC in $CD25^+$ HEK293T cells over-expressing wild type or mutant Tet2 (see **Suppl**. **Fig 3**) using the anti-5-hmC (**c**) or anti-5-mC (**d**) antibody as described above.

Suppl. Figure 7. 5-hmC (CMS) levels in genomic DNA samples from healthy donors and patients with different haematological malignancies. 5-hmC levels were assessed by dot blot assay of bisulfite-treated DNAs, serially diluted and probed with the anti-CMS antiserum.

a, Relative 5-hmC (CMS) levels from experiments (Expts.) 2-4 are shown, normalised to the average 5-hmC level in patients with wild type *TET2* (set at 1) (for patient data, see **Suppl. Table 1**). The horizontal bars indicate the mean.

b, 5-hmC (CMS) levels in genomic DNA from healthy donors and patients with haematological malignancies. Data were plotted as median ± S.D. ($n=3$). Black bars: healthy donors, blue bars: MDS, orange bars: MPN, red bars: MDS/ MPN, purple bars: AML, green bars: sAML derived from the above conditions.

Suppl. Figure 8. A representative scheme for haematopoietic cell sorting. Total enucleated bone marrow cells **(a-d)** or thymocytes **(e)** were stained with antibodies as described and haematopoietic subsets were sorted using the following criteria.

a, LSK ($Lin^-Sca\text{-}1^+c\text{-}Kit^+IL\text{-}7R\alpha^-$) and CLP ($Lin^-IL\text{-}7R\alpha^+c\text{-}Kit^{lo}Sca\text{-}1^{lo}$).

b, CMP (Lin$^-$Sca-1$^-$c-Kit^{hi}CD34$^+$FcγRII/III^{lo}), GMP (Lin$^-$Sca-1$^-$c-Kit^{hi}CD34$^+$FcγRII/III^{hi}) and MEP (Lin$^-$Sca-1$^-$c-Kit^{hi}CD34$^-$FcγRII/III^{lo}).

c, granulocytes (Gr-1$^+$Mac-1$^+$), monocytes (Gr-1$^-$Mac-1$^+$) and early erythroblasts (Ter-119$^+$).

d, pro-B (B220$^+$IgM$^-$CD43$^+$), pre-B (B220$^+$IgM$^-$CD43$^-$), immature (B220$^+$IgM$^+$) and mature (B220^{hi}IgM$^+$) B cells.

e, DN1 (CD4$^-$CD8$^-$CD44$^+$CD25$^-$), DN2 (CD4$^-$CD8$^-$CD44$^+$CD25$^+$), DN3 (CD4$^-$CD8$^-$CD44$^-$CD25$^+$), DN4 (CD4$^-$CD8$^-$CD44$^-$CD25$^-$), ISP (CD4$^-$CD8$^+$TCRβ$^{-/lo}$), CD4$^+$ SP (CD4$^+$CD8$^-$TCRβ$^+$) and CD8$^+$ SP (CD4$^-$CD8$^+$TCRβ$^+$).

Suppl. Figure 9. Expression of Tet family members and 5-hmC levels in haematopoietic subsets.

a, *Tet1*, *Tet2* and *Tet3* mRNA expression during haematopoiesis. Cells along the myeloid and lymphoid (B and T cell) differentiation pathways were isolated from bone marrow and thymus of C57BL/6 mice by flow cytometry (**Suppl. Fig. 8**), and quantitative RT-PCR analysis was performed to assess the level of *Tet1, Tet2* and *Tet3* mRNA (shown after normalisation to the level of *Gapdh* mRNA in the same cell population). Total RNA from undifferentiated ES cells and ES cells treated with 1 μM retinoic acid for 4 days was also analyzed for comparison. The relative normalized levels of *Tet1-3* mRNA are shown, with the amount in the LSK population arbitrarily set to 1.

b-c, 5-hmC levels in bone marrow (**b**) and thymus (**c**) cell populations were assessed by quantitative dot blot assay with anti-CMS antibody after treatment of genomic DNA with bisulfite.

Suppl. Figure 10. *In vitro* differentiation in the presence of low cytokine concentrations in control and Tet2-depleted cells.

a, Schematic representation of retroviral constructs for RNAi-mediated knockdown of Tet2. Shown are control (LMP) and Tet2-KD retroviral vectors, encoding either GFP alone, or both GFP and shRNA against *Tet2* mRNA, respectively. The scramble retroviral vector (not shown) contains unrelated sequences instead of sh*Tet2* sequences.

b, Use of *Tet2* shRNA #3 to suppress *Tet2* gene expression in bone marrow cells. When the four SMARTpool (Dharmacon) siRNAs were tested individually in V6.5 ES cells, siRNAs #1 and #3 gave equivalent efficiency of knockdown. However, shRNA #3 gave better knockdown than shRNA #1 when cloned into the vector **pSUPER-puro** in V6.5 ES cells (K.P.K. and A.R. unpublished data). We therefore tested siRNAs #1 and #3 in the **LMP** retroviral vector by transducing c-Kit$^+$ bone marrow cells (Left) and NIH3T3 cells (Right). Only one of the two shRNA sequences (#3) was effective in bone marrow cells and NIH3T3 cells. Error bars show the range of duplicates.

c, c-Kit$^+$ bone marrow stem/progenitor cells were magnetically sorted from bone marrow of C57BL/6 mice and transduced as described. After selection with puromycin (2 μg/ml) for 3 days, GFP expression was analyzed by flow cytometry. Cells were subject to quantitative RT-PCR (see **Fig. 3a).**

d, Lin$^-$ cells were transduced with control (scramble) or shTet2 retroviruses as shown in **Fig. 3c**, then

cultured in the presence of 50 ng/ml stem cell factor (SCF), puromycin (2 μg/ml) and cytokines at 1 ng/ml as indicated. After 4 days, flow cytometric analysis of Mac-1 *vs.* F4/80 (*left panel*) or CD115 (also known as M-CSFR) (*right panel*) was performed. All cells were GFP^+ at the day of analysis.

Suppl. Figure 11. Tet2 deficiency does not affect the acute proliferation of developing myeloid cells.

a, Lin^- cells were transduced with control (scramble) or shTet2 retroviruses and then cultured in the presence of 50 ng/ml SCF, puromycin (2 μg/ml) and 10 ng/ml GM-CSF. After 4 days, cells were pulse-labeled with 10 μM BrdU (BD Biosciences) for 1 hr and stained to determine cell surface CD115 (M-CSFR) expression and BrdU incorporation.

b, BrdU incorporation was assessed after gating on $CD115^+$ ($M\text{-}CSFR^+$) monocyte/macrophage cells.