Fig. 2. Two-dimensional TLC, HPLC, and MS identification of hmC. (**A**) Two-dimensional TLC analysis of synthetic DNA templates indicates that hmC comigrates with the "x" spot (Fig. 1). (**B**) HPLC chromatograms (A, 254 nm) of the nucleosides derived from synthetic and cerebellum DNA. The peaks were identified by MS. The arrow points to the peak, which elutes at the same time as hmdC. (**C**) MS of the fraction corresponding to the HPLC peak indicated above. Closed arrows indicate the masses of 5-hydroxymethylcytosine and 5-hydroxymethyl-2'-deoxycytidine sodium ions (structures are shown in the insets). Open arrows indicate the ions generated by 2'-deoxycytidine, which elutes in a large nearby peak and spills over into the analyzed fraction.

References and Notes

1. J. P. Doyle et al., Cell **135**, 749 (2008).
2. M. Heiman et al., Cell **135**, 738 (2008).
3. G. R. Wyatt, S. S. Cohen, Biochem. J. **55**, 774 (1953).
4. S. Tardy-Planechaud, J. Fujimoto, S. S. Lin, L. C. Sowers, Nucleic Acids Res. **25**, 553 (1997).
5. A. Burdzy, K. T. Noyes, V. Valinluck, L. C. Sowers, Nucleic Acids Res. **30**, 4068 (2002).
6. S. Zuo, R. J. Boorstein, G. W. Teebor, Nucleic Acids Res. **23**, 3239 (1995).
7. J. Cadet, T. Douki, J. L. Ravanat, Nat. Chem. Biol. **2**, 348 (2006).
8. N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, Biochem. J. **126**, 781 (1972).
9. R. M. Kothari, V. Shankar, J. Mol. Evol. **7**, 325 (1976).
10. J. H. Gommers-Ampt, P. Borst, FASEB J. **9**, 1034 (1995).
11. H. Hayatsu, M. Shiragami, Biochemistry **18**, 632 (1979).
12. S. K. Ooi, T. H. Bestor, Cell **133**, 1145 (2008).
13. V. Valinluck et al., Nucleic Acids Res. **32**, 4100 (2004).
14. B. H. Ramsahoye, Methods Mol. Biol. **200**, 9 (2002).
15. We thank B. Gauthier for technical assistance; S. Mazel, C. Bare, and X. Fan for flow cytometry advice and nuclei sorts; and H. Deng and J. Fernandez for acquirement of MS data and help with HPLC. We are grateful to members of the Heintz laboratory for discussions and support. This work was supported by the Howard Hughes Medical Institute and the Simons Foundation Autism Research Initiative.

Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1

Mamta Tahiliani,[1] Kian Peng Koh,[1] Yinghua Shen,[2] William A. Pastor,[1] Hozefa Bandukwala,[1] Yevgeny Brudno,[2] Suneet Agarwal,[3] Lakshminarayan M. Iyer,[4] David R. Liu,[2]* L. Aravind,[4]* Anjana Rao[1]*

DNA cytosine methylation is crucial for retrotransposon silencing and mammalian development. In a computational search for enzymes that could modify 5-methylcytosine (5mC), we identified TET proteins as mammalian homologs of the trypanosome proteins JBP1 and JBP2, which have been proposed to oxidize the 5-methyl group of thymine. We show here that TET1, a fusion partner of the MLL gene in acute myeloid leukemia, is a 2-oxoglutarate (2OG)- and Fe(II)-dependent enzyme that catalyzes conversion of 5mC to 5-hydroxymethylcytosine (hmC) in cultured cells and in vitro. hmC is present in the genome of mouse embryonic stem cells, and hmC levels decrease upon RNA interference–mediated depletion of TET1. Thus, TET proteins have potential roles in epigenetic regulation through modification of 5mC to hmC.

5-methylcytosine (5mC) is a minor base in mammalian DNA: It constitutes ~1% of all DNA bases and is found almost exclusively as symmetrical methylation of the dinucleotide CpG (1). The majority of methylated CpG is found in repetitive DNA elements, suggesting that cytosine methylation evolved as a defense against transposons and other parasitic elements (2). Methylation patterns change dynamically in early embryogenesis, when CpG methylation is essential for X-inactivation and asymmetric expression of imprinted genes (3). In somatic cells, promoter methylation often shows a correlation with gene expression: CpG methylation may directly interfere with the binding of certain transcriptional regulators to their cognate DNA sequences or may enable recruitment of methyl-CpG binding proteins that create a repressed chromatin environment (4). DNA methylation patterns are highly dysregulated in cancer: Changes in methylation status have been postulated to inactivate tumor suppressors and activate oncogenes, thus contributing to tumorigenesis (5).

[1]Department of Pathology, Harvard Medical School and Immune Disease Institute, 200 Longwood Avenue, Boston, MA 02115, USA. [2]Department of Chemistry and Chemical Biology and the Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA. [3]Division of Pediatric Hematology/Oncology, Children's Hospital Boston and Dana-Farber Cancer Institute, Boston, MA 02115, USA. [4]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

*To whom correspondence should be addressed. E-mail: arao@idi.harvard.edu (A.R.); aravind@ncbi.nlm.nih.gov (L.A.); drliu@fas.harvard.edu (D.R.L.)

Fig. 1. Expression of TET1 in HEK293 cells results in decreased 5mC staining. (**A**) Predicted domain architecture of human TET1 showing the CXXC-type zinc-binding domain (amino acid 584 to 624); the cysteine-rich region (Cys-rich) (amino acid 1418 to 1610); the DSBH domain (amino acid 1611 to 2074); and three bipartite nuclear localization sequences (NLS). (**B**) HEK293 cells overexpressing wild-type or mutant HA-TET1 were costained with antibodies specific for the HA epitope (green) and 5mC (red). To orient the reader, some HA-expressing cells are circled. Scale bar, 10 μm. (**C**) Staining intensities of HA and 5mC were measured in individual nuclei. Data from HA-TET1–expressing cell populations are presented as dot plots (red), superimposed on dot plots from mock-transfected cells (blue), with each dot representing an individual cell. (**D**) Quantification of 5mC staining intensity of HA-positive cells compared with that of mock-transfected cells (set to 1). Data shown are mean ± SEM and are representative of three experiments.
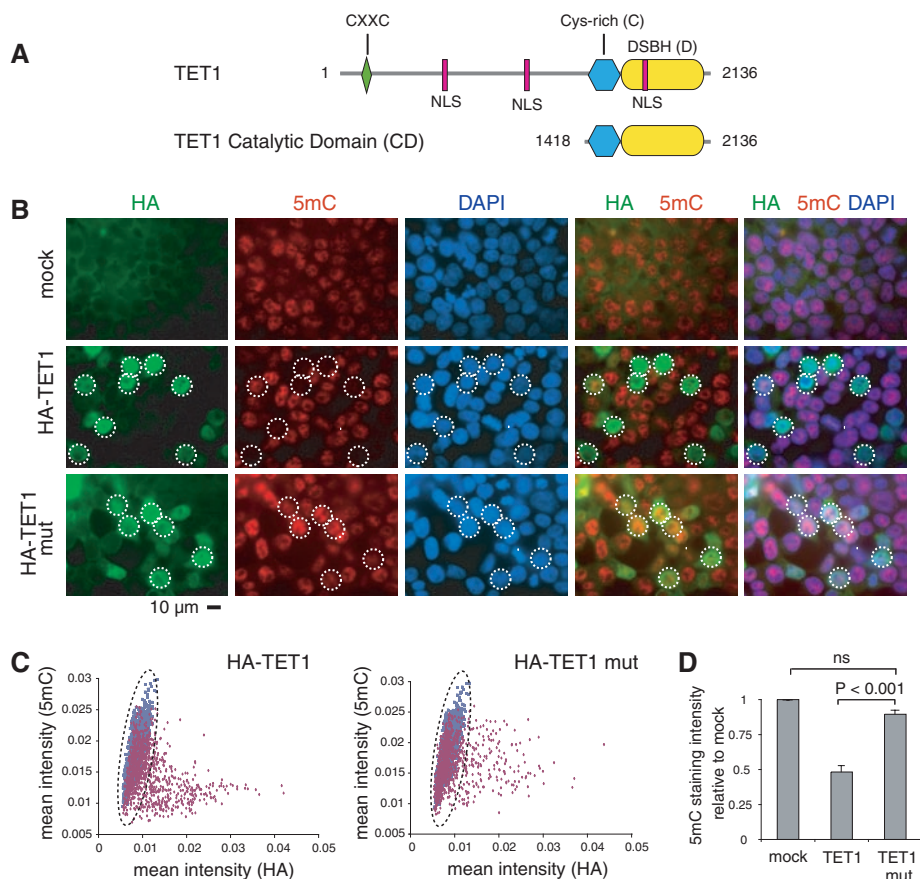
Fig. 2. Genomic DNA of TET1-overexpressing cells contains a modified nucleotide within the dinucleotide CG. (**A**) Description of experimental design. (**B** to **F**) Genomic DNA was purified from HEK293 cells overexpressing TET1 and cleaved with [(**B**) and (**C**)] MspI, (**D**) HpaII, or [(**E**) and (**F**)] TaqᵅI. The fragments were end-labeled, digested to 5′ dNMPs, and resolved by TLC. A modified nucleotide (subsequently identified as hm-dCMP) is indicated by ?. Neither 5m-dCMP nor the modified nucleotide are observed when the DNA is digested with HpaII. [(**C**) and (**F**)] Quantification of the relative abundance of dCMP, 5m-dCMP, and the modified nucleotide. Data shown are mean ± SD of three independent transfections and are representative of at least three experiments.
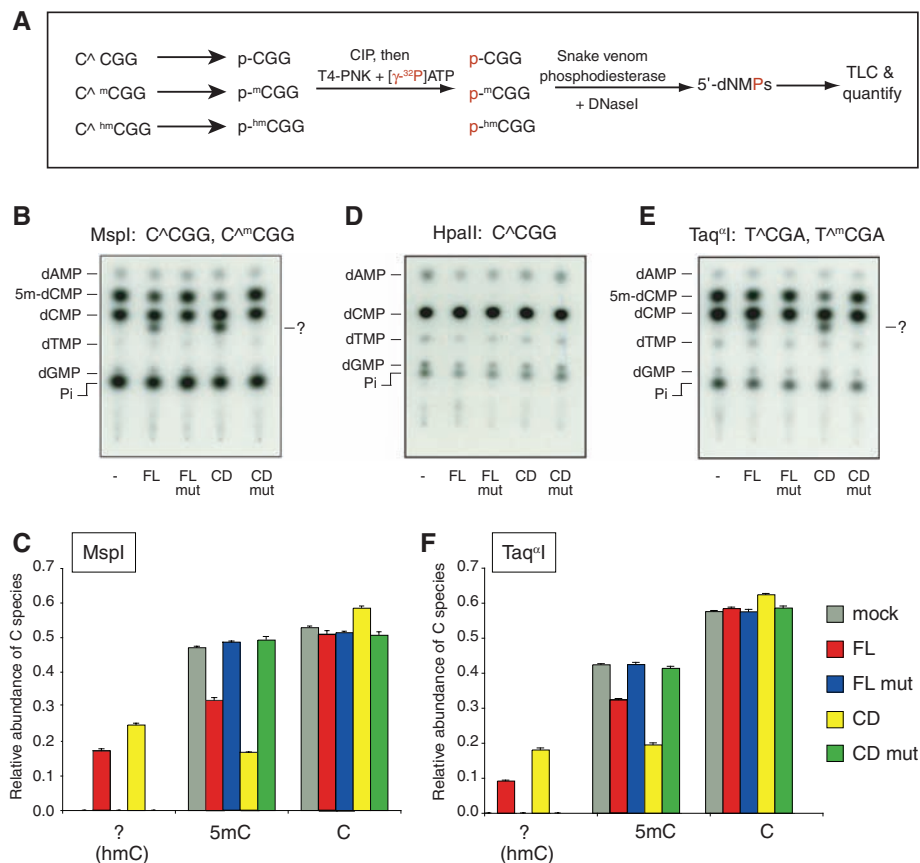
Trypanosomes contain base J (β-D-glucosyl hydroxymethyluracil), a modified thymine produced by sequential hydroxylation and glucosylation of the methyl group of thymine (fig. S1A) (6). J biosynthesis requires JBP1 and JBP2, enzymes of the 2OG- and Fe(II)-dependent oxygenase superfamily predicted to catalyze the first step of J biosynthesis (7, 8). Like 5-methylcytosine, base J has an association with gene silencing: It is present in silenced copies of the genes encoding the variable surface glycoprotein (VSG) responsible for antigenic variation in the host but is absent from the single expressed copy (6). We performed a computational search for homologs of JBP1 and JBP2 in the hope of identifying mammalian enzymes that modified 5mC.

Iterative sequence profile searches using the predicted oxygenase domains of JBP1 and JBP2 recovered homologous regions in three paralogous human proteins TET1, TET2, and TET3 and their orthologs found throughout metazoa ($e < 10^{-5}$), as well as homologous domains in fungi and algae (fig. S2 and SOM Text). Secondary structure predictions suggested the existence of an N-terminal α helix followed by a continuous series of β strands, typical of the double-stranded β helix (DSBH) fold of the 2OG-Fe(II) oxygenases (fig. S3) (9). A multiple sequence alignment showed that the new TET/JBP family displayed all of the typical features of 2OG-Fe(II) oxygenases, including conservation of residues predicted to be important for coordination of the cofactors Fe(II) and 2OG (fig. S3 and SOM Text). The metazoan TET proteins contain a unique conserved cysteine-rich region, contiguous with the N terminus of the DSBH region (Fig. 1A and SOM Text). Vertebrate TET1 and TET3, and their orthologs from all other animals, also possess a CXXC domain, a binuclear Zn-chelating domain, found in several chromatin-associated proteins, that in certain cases has been shown to discriminate between methylated and unmethylated DNA (10).

As a first step in determining whether TET proteins operate on 5mC, we transfected human embryonic kidney (HEK) 293 cells with full-length hemagglutinin (HA)–tagged TET1, then stained the transfected cells for 5mC and the HA epitope. Mock-transfected cells showed substantial variation in 5mC staining intensity (fig. S4; Fig. 1B, top panel; quantified in Fig. 1C), either because 5mC levels vary from cell to cell or because the accessibility of 5mC to the antibody differs among cells because of technical considerations (e.g., incomplete denaturation of DNA). Cells transfected with wild-type TET1 showed a strong correlation of HA positivity with decreased staining for 5mC, both visually (Fig. 1B, middle panel) and by quantification (Fig. 1C, left panel, and Fig. 1D). Untransfected HA-low cells showed a spread of 5mC staining intensity similar to that of mock-transfected cells (Fig. 1C, left panel; note overlapping red and blue dots at low HA intensities), whereas productively transfected HA-high cells showed uniformly low 5mC staining intensity (Fig. 1C, left panel, red HA-high dots). Cells transfected with mutant

TET1 bearing H1671Y, D1673A substitutions predicted to impair Fe(II) binding did not show decreased staining for 5mC (Fig. 1B, bottom panel; Fig. 1C, right panel; and Fig. 1D).

To determine quantitatively whether TET1 overexpression affects intracellular 5mC levels, we measured the ratio of 5mC to C at a subset of genomic CpG sites in cells expressing either full-length TET1 (TET1-FL) or the predicted catalytic domain of TET1 [TET1-CD, comprising the Cys-rich (C) and DSBH (D) regions; Fig. 1A]. HEK293 cells were transiently transfected with plasmids in which TET1 expression was coupled to expression of human CD25 from an internal ribosome

entry site (IRES). Genomic DNA from CD25-expressing cells was digested with MspI, which cleaves DNA at the sequence C^CGG regardless of whether the second C is methylated. The resulting fragments, whose 5′ ends derive from the dinucleotide CpG and contain either C or 5mC, were end-labeled and digested to yield 5′ phosphorylated dNMPs that were resolved using thin-layer chromatography (TLC) (Fig. 2A) (11).

MspI-digested DNA from cells transfected with the control vector yielded predominantly dCMP and 5m-dCMP as expected (Fig. 2B, lane 1), whereas DNA from cells expressing wild-type TET1-FL or TET1-CD yielded an additional
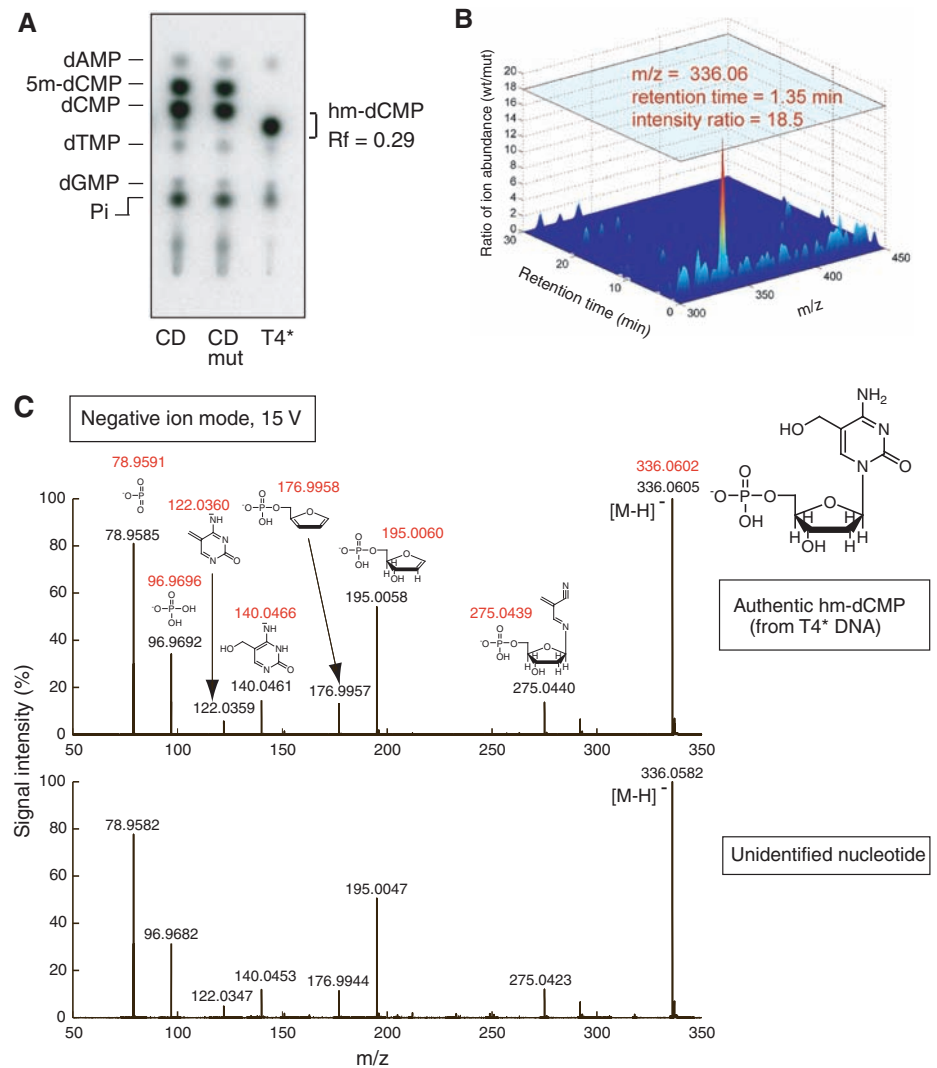
Fig. 3. The modified nucleotide is identified as 5-hydroxymethylcytosine. (A) Genomic DNA from T4 phage grown in UDP-glucose–deficient E. coli ER1656 (T4*) and HEK293 cells transfected with wild-type or mutant TET1-CD were digested with Taq$^\alpha$I. The fragments were end-labeled, digested to mononucleotides, and separated by TLC. The modified nucleotide present in TET1-CD–expressing cells migrates similarly to authentic hm-dCMP. (B) Comparison of liquid chromatography–electrospray ionization MS ions present at an Rf = 0.29 in genomic DNA from cells expressing wild-type (wt) or mutant (mut) TET1-CD. A species with an observed m/z = 336.06 was more abundant by a factor of 18.5 in the wild-type sample compared with the mutant sample. (C) Mass spectrometry fragmentation (MS/MS) analysis of authentic hm-dCMP (top), and the m/z = 336.06 species isolated from cells expressing TET1-CD (bottom). Expected m/z values are shown in red; observed m/z values are shown in black (anticipated mass accuracy is within 0.002 Da).
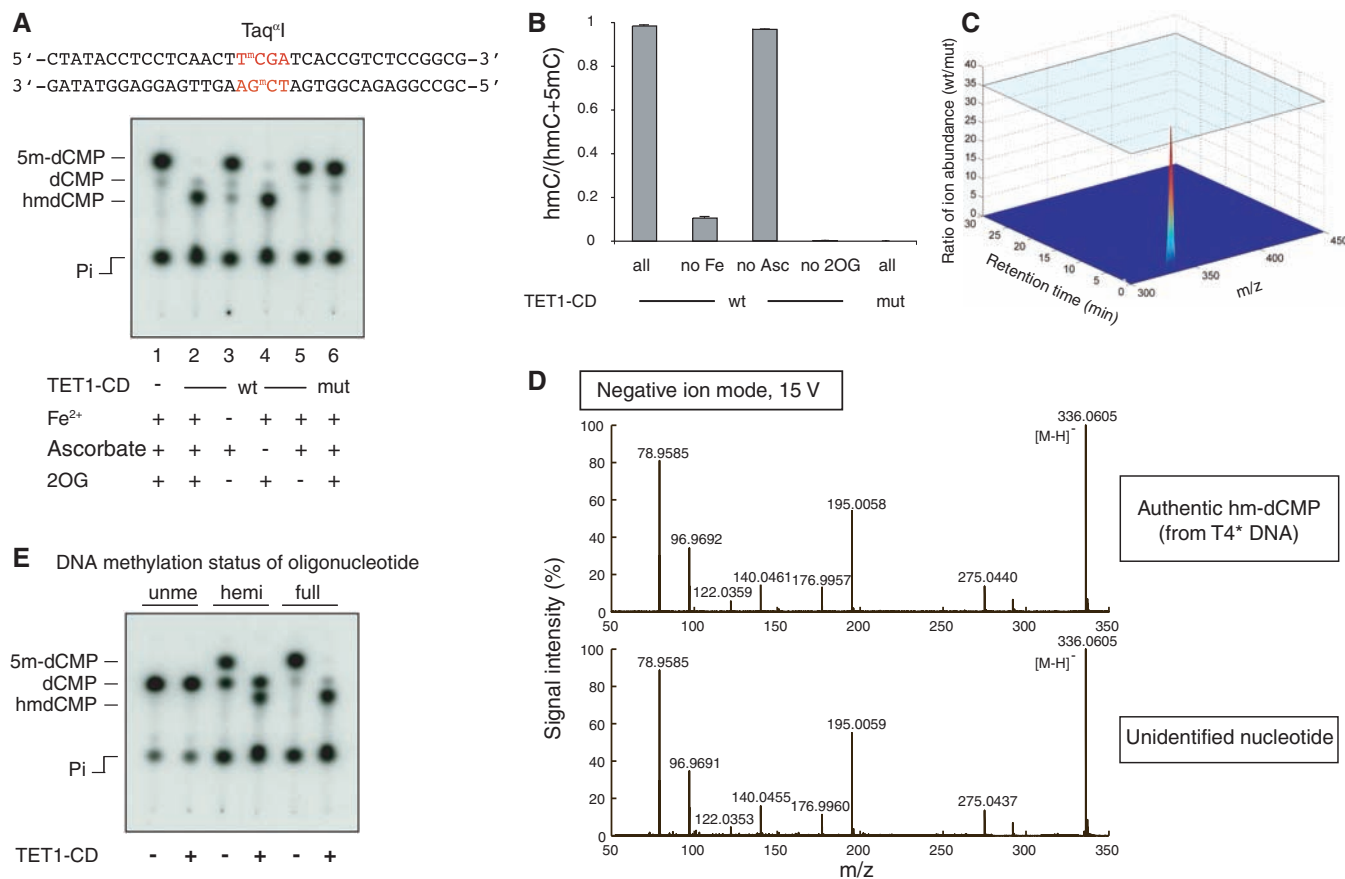
Fig. 4. Recombinant Flag-HA-TET1-CD purified from Sf9 cells converts 5mC to hmC in methylated DNA oligonucleotides in vitro. (**A**) Double-stranded DNA oligonucleotides containing a fully methylated $Taq^{\alpha}I$ site were incubated with wt or mut Flag-HA-TET1-CD (1:10 enzyme to substrate ratio). Recovered oligonucleotides were digested with $Taq^{\alpha}I$ and analyzed by TLC. The faint dCMP spot in each lane is derived from end-labeling of the C at the 5′ end of each strand of the substrate. (**B**) The extent of conversion of 5mC to hmC is shown as the mean ratio [hmC/(hmC+ 5mC)] $\pm$ SD. (**C**) Comparison of species at an Rf of 0.29 in products resulting from incubation with wt or mut Flag-HA-TET1-CD. (**D**) MS fragmentation analysis of authentic hm-dCMP (top) and the nucleotide generated by Flag-HA-TET1-CD (bottom). Observed masses are shown in black (mass accuracy was within 0.002 Da). (**E**) Recombinant Flag-HA-TET1-CD is able to hydroxylate 5mC in fully methylated (full, lanes 5 and 6) and hemi-methylated (hemi, lanes 3 and 4) substrates. Unme, unmethylated DNA oligonucleotide (lanes 1 and 2).

unidentified labeled species migrating more slowly than dCMP ("?," Fig. 2B, lanes 2 and 4). This new species was not detected when the DNA was digested with HpaII, the methylation-sensitive isoschizomer of MspI (Fig. 2D) but was clearly observed when the DNA was digested with $Taq^{\alpha}I$, a methylation-insensitive enzyme that cuts at a different sequence, T^CGA (Fig. 2E). The new species was not observed in MspI- or $Taq^{\alpha}I$-digested DNA from cells transfected with mutant TET1-FL or TET1-CD (Fig. 2, B and E, lanes 3 and 5), and its appearance was associated with decreased abundance of 5m-dCMP (Fig. 2, B and E, lanes 2 and 4; quantified in Fig. 2, C and F). In all experiments, expression of wild-type TET1-CD correlated with a small but significant increase in the abundance of dCMP (Fig. 2, C and F). Together these results suggest that the unidentified species is derived through modification of 5m-dCMP and may be an intermediate in the passive or active conversion of 5mC to C.

We used high-resolution mass spectrometry (MS) to identify the novel nucleotide. Genomic DNA was prepared from HEK293 cells overexpressing wild-type or mutant TET1-CD. A singly charged species with an observed mass/charge ratio (m/z) of 336.0582, consistent with a molecular formula of $C_{10}H_{15}NO_8P^-$, was the only species migrating at the expected position on TLC that exhibited a large (by a factor of about 19) difference in abundance between the wild-type and mutant samples (Fig. 3, A and B). Based on this result, our computational analyses, and the data of Fig. 2, we hypothesized that the unidentified species was 5-hydroxymethylcytosine (hmC), produced by TET1 through hydroxylation of the methyl group of 5mC (fig. S1). As a standard, we prepared authentic hm-dCMP from unglucosylated DNA of T4 phage grown in *Escherichia coli* ER1656, a strain deficient in the glucose donor molecule UDP-glucose (abbreviated T4* DNA) (*12*). TLC assays showed that the novel nucleotide generated in unsorted cells expressing TET1-CD migrated similarly to hm-dCMP from T4* DNA (Fig. 3A). Tandem mass spectrometry (MS-MS) fragmentation experiments at several collision energies (15 and 25 V) in both positive and negative ion modes confirmed that the fragmentation pattern of the 336.0582 dalton ion was identical to that of hm-dCMP (Fig. 3C and fig. S5).

To establish that TET1 was directly responsible for hmC production, we expressed Flag-HA-tagged wild-type and mutant TET1-CD in Sf9 insect cells, purified the recombinant proteins to near homogeneity (fig. S6A), and assayed their catalytic activity on fully methylated double-stranded DNA oligonucleotides. Wild-type, but not mutant, TET1-CD catalyzed robust conversion of 5mC to hmC and displayed an absolute requirement for both Fe(II) and 2OG (Fig. 4A; quantified in Fig. 4B). Omission of ascorbate did not result in a significant decrease in catalytic activity, most likely because we included dithiothreitol in the reaction to counteract the strong tendency of TET1-CD to oxidize (fig. S6, A to D) (*13–15*). Recombinant TET1-CD was specific for 5mC: We did not detect conversion of thymine to hydromethyluracil (hmU) (fig. S7).

We used high-resolution MS to demonstrate, as before, that a singly charged species at m/z of 336.0582 was the only species migrating at the expected position that differed significantly (by a

Fig. 5. hmC is present in ES cell DNA, and its abundance decreases upon differentiation or Tet1 depletion. (**A**) (Left) TLC showing that hmC is detected in the genome of undifferentiated ES cells but not dendritic cells or T cells. (Right) Quantification of the relative abundance by PhosphorImager of 5mC, C, and hmC in the genomic DNA of undifferentiated ES cells. (**B**) Tet1 mRNA levels decline by ~80% in ES cells induced to differentiate by withdrawal of LIF for 5 days. (**C**) The same differentiated ES cells show ~40% decrease in hmC levels. (**D**) Transfection of ES cells using two different RNAi duplexes directed against Tet1 decreases Tet1 mRNA levels by ~75%. (**E**) The same Tet1-depleted ES cells show ~40% decline in hmC levels. Data shown are mean ± SD and are representative of 2 to 3 experiments.
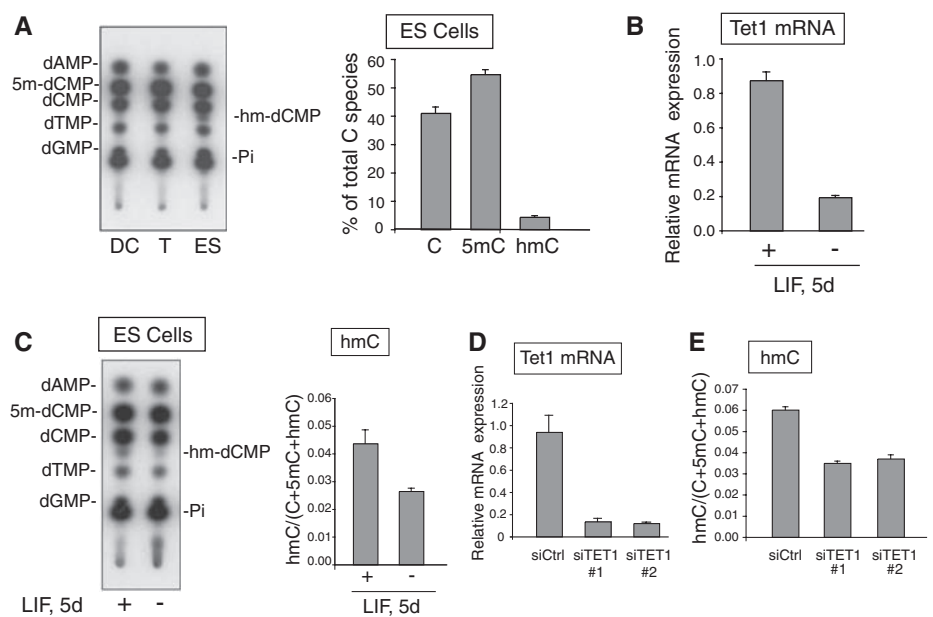
factor of 35) in abundance when comparing substrates incubated with wild-type and mutant proteins (Fig. 4C). MS-MS experiments confirmed that the fragmentation pattern of the species produced by recombinant TET1-CD was identical to that of authentic hm-dCMP in unglucosylated T4 DNA (Fig. 4D). TET1-CD was also able to oxidize 5mC to hmC in hemimethylated double-stranded DNA (Fig. 4E).

We asked whether hmC was a physiological constituent of mammalian DNA. Using the TLC assay, we observed a clear spot corresponding to labeled hmC in mouse embryonic stem (ES) cells but not in previously activated human T cells or mouse dendritic cells (Fig. 5A). Quantification of multiple experiments indicated that hmC and 5mC constituted 4 to 6% and 55 to 60%, respectively, of all cytosine species in MspI cleavage sites (C^CGG) in ES cells (Fig. 5A). Tet1 mRNA levels declined by 80% in response to leukemia inhibitory factor (LIF) withdrawal for 5 days, compared with the levels observed in undifferentiated ES cells (Fig. 5B); in parallel, hmC levels diminished from 4.4 to 2.6% of total C species, a decline of ~40% from control levels (Fig. 5C). The difference might be due to the compensatory activity of other Tet-family proteins. Similarly, RNA interference (RNAi)–mediated depletion of endogenous Tet1 resulted in an 87% decrease in Tet1 mRNA levels and a parallel ~40% decrease in hmC levels (Fig. 5, D and E). Again, the difference is likely due to the presence of Tet2 and Tet3, which are both expressed in ES cells.

Together these data strongly support the hypothesis that Tet1, and potentially other Tet family members, are responsible for hmC generation in ES cells under physiological conditions (fig. S8A). CpG dinucleotides are ~0.8% of all dinucleotides in the mouse genome (16); thus, hmC (which constitutes ~4% of all cytosine species in CpG dinucleotides located in MspI cleavage sites) is

~0.032% of all bases (~1 in every 3000 nucleotides, or ~2 × 10⁶ bases per haploid genome). For comparison, 5mC is 55 to 60% of all cytosines in CpG dinucleotides in MspI cleavage sites (Fig. 5A), about 14 times as high as hmC (hmC may not be confined to CpG) (SOM Text). An important question is whether hmC and TET proteins are localized to specific regions of ES cell DNA—for instance, genes that are involved in maintaining pluripotency or that are poised to be expressed upon differentiation. A full appreciation of the biological importance of hmC will require the development of tools that allow hmC, 5mC, and C to be distinguished unequivocally (SOM Text).

As a potentially stable base (SOM Text), hmC may influence chromatin structure and local transcriptional activity by recruiting selective hmC-binding proteins or excluding methyl-CpG–binding proteins (MBPs) that normally recognize 5mC, thus displacing chromatin-modifying complexes recruited by MBPs (fig. S8B, center). Indeed, it has already been demonstrated that the methyl-binding protein MeCP2 does not recognize hmC (17). Alternatively, conversion of 5mC to hmC may facilitate passive DNA demethylation by excluding the maintenance DNA methyltransferase DNMT1, which recognizes hmC poorly (fig. S8B, left) (18). Even a minor reduction in the fidelity of maintenance methylation would be expected to result in an exponential decrease in CpG methylation over the course of many cell cycles. Finally, hmC may be an intermediate in a pathway of active DNA demethylation (fig. S8B, right). hmC has been shown to yield cytosine through loss of formaldehyde in photooxidation experiments (19) and at high pH (20, 21), leaving open the possibility that hmC could convert to cytosine under certain conditions in cells. A related possibility is that specific DNA repair mechanisms replace hmC or its derivatives with C (22, 23). In support of this hypothesis, a glycosylase activity specific for hmC was reported

in bovine thymus extracts (24). Moreover, several DNA glycosylases, including TDG and MBD4, have been implicated in DNA demethylation, although none of them has shown convincing activity on 5mC in in vitro enzymatic assays (25–27). Cytosine deamination has also been implicated in demethylation of DNA (26–28); in this context, deamination of hmC yields hmU, and high levels of hmU:G glycosylase activity have been reported in fibroblast extracts (29).

These studies alter our perception of how cytosine methylation may be regulated in mammalian cells. Notably, disruptions of the TET1 and TET2 genetic loci have been reported in association with hematologic malignancies (SOM Text). A fusion of TET1 with the histone methyltransferase MLL has been identified in several cases of acute myeloid leukemia (AML) associated with t(10;11)(q22;q23) translocation (30, 31). Homozygous null mutations and chromosomal deletions involving the TET2 locus have been found in myeloproliferative disorders, suggesting a tumor suppressor function for TET2 (32, 33). It will be important to test the involvement of TET proteins and hmC in oncogenic transformation and malignant progression.

References and Notes
1. M. Ehrlich, R. Y. Wang, *Science* **212**, 1350 (1981).
2. M. G. Goll, T. H. Bestor, *Annu. Rev. Biochem.* **74**, 481 (2005).
3. W. Reik, *Nature* **447**, 425 (2007).
4. A. Bird, *Genes Dev.* **16**, 6 (2002).
5. E. N. Gal-Yam, Y. Saito, G. Egger, P. A. Jones, *Annu. Rev. Med.* **59**, 267 (2008).
6. P. Borst, R. Sabatini, *Annu. Rev. Microbiol.* **62**, 235 (2008).
7. Z. Yu *et al.*, *Nucleic Acids Res.* **35**, 2107 (2007).
8. L. J. Cliffe *et al.*, *Nucleic Acids Res.* **37**, 1452 (2009).
9. L. Aravind, E. V. Koonin, *Genome Biol.* **2**, RESEARCH0007 (2001).
10. M. D. Allen *et al.*, *EMBO J.* **25**, 4503 (2006).
11. H. Cedar, A. Solage, G. Glaser, A. Razin, *Nucleic Acids Res.* **6**, 2125 (1979).
12. G. R. Wyatt, S. S. Cohen, *Biochem. J.* **55**, 774 (1953).

13. L. Que Jr., R. Y. Ho, *Chem. Rev.* **96**, 2607 (1996).
14. C. Loenarz, C. J. Schofield, *Nat. Chem. Biol.* **4**, 152 (2008).
15. L. E. Netto, E. R. Stadtman, *Arch. Biochem. Biophys.* **333**, 233 (1996).
16. M. W. Simmen, *Genomics* **92**, 33 (2008).
17. V. Valinluck *et al.*, *Nucleic Acids Res.* **32**, 4100 (2004).
18. V. Valinluck, L. C. Sowers, *Cancer Res.* **67**, 946 (2007).
19. E. Privat, L. C. Sowers, *Chem. Res. Toxicol.* **9**, 745 (1996).
20. J. G. Flaks, S. S. Cohen, *J. Biol. Chem.* **234**, 1501 (1959).
21. A. H. Alegria, *Biochim. Biophys. Acta* **149**, 317 (1967).
22. S. K. Ooi, T. H. Bestor, *Cell* **133**, 1145 (2008).
23. J. Jiricny, M. Menigatti, *Cell* **135**, 1167 (2008).
24. S. V. Cannon, A. Cummings, G. W. Teebor, *Biochem. Biophys. Res. Commun.* **151**, 1173 (1988).
25. B. Zhu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5135 (2000).
26. R. Metivier *et al.*, *Nature* **452**, 45 (2008).
27. S. Kangaspeska *et al.*, *Nature* **452**, 112 (2008).
28. K. Rai *et al.*, *Cell* **135**, 1201 (2008).
29. V. Rusmintratip, L. C. Sowers, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14183 (2000).
30. R. Ono *et al.*, *Cancer Res.* **62**, 4075 (2002).
31. R. B. Lorsbach *et al.*, *Leukemia* **17**, 637 (2003).
32. F. Viguie *et al.*, *Leukemia* **19**, 1411 (2005).
33. F. Delhommeau *et al.*, paper presented at the American Society of Hematology Annual Meeting and Exposition, San Francisco, CA, December 9, 2008.
34. We thank K. Kreuzer for the gift of the T4 phage and *E. coli* ER1656 and CR63; Charles Richardson, Udi Qimron, and Ben Beauchamp for advice and assistance in culturing T4 phage; Melissa Call for advice on production of recombinant proteins in insect cells; and Patrick Hogan for many helpful discussions. This work was supported by NIH grant AI44432, a Scholar Award from the Juvenile Diabetes Research Foundation, and a Seed grant from the Harvard Stem Cell Institute (to A.R.); Howard Hughes Medical Institute and NIH/NIGMS

A Functional Role for Transposases in a Large Eukaryotic Genome

Mariusz Nowacki,[1] Brian P. Higgins,[1] Genevieve M. Maquilan,[1] Estienne C. Swart,[1] Thomas G. Doak,[2] Laura F. Landweber[1]*

Despite comprising much of the eukaryotic genome, few transposons are active, and they usually confer no benefit to the host. Through an exaggerated process of genome rearrangement, *Oxytricha trifallax* destroys 95% of its germline genome during development. This includes the elimination of all transposon DNA. We show that germline-limited transposase genes play key roles in this process of genome-wide DNA excision, which suggests that transposases function in large eukaryotic genomes containing thousands of active transposons. We show that transposase gene expression occurs during germline-soma differentiation and that silencing of transposase by RNA interference leads to abnormal DNA rearrangement in the offspring. This study suggests a new important role in *Oxytricha* for this large portion of genomic DNA that was previously thought of as junk.

*O*xytricha trifallax, a ciliated protozoan, is a unicellular eukaryote with two types of nuclei in the same cytoplasm. Diploid micronuclei are transcriptionally silent during vege- tative growth but transmit the germline genome through sexual conjugation (fig. S1). After sex, the gene-rich macronuclei that govern somatic gene expression degrade, and new macronuclei develop from germline micronuclei. Macronuclear development in *Oxytricha* (also called *Sterkiella*) involves massive chromosome fragmentation and the deletion of thousands of copies of germline transposons (selfish DNA elements that are capa- ble of self-replication in their host genomes) plus the destruction of intergenic DNA and noncoding DNA sequences that interrupt genes [internally eliminated sequences (IESs)]. This leads to a 95% reduction in sequence complexity, compressing the 1-Gb germline genome into a gene-dense macronuclear genome of only ~50 Mb. Sorting and reordering the hundreds of thousands of re- maining short pieces [macronuclear-destined seg- ments (MDSs)] produces mature macronuclear sequences. The resulting ~2-kb nanochromo-

[1]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA. [2]Department of Biol- ogy, Indiana University, Bloomington, IN 47405, USA.

*To whom correspondence should be addressed. E-mail: lfl@princeton.edu

Fig. 1. Expression of TBE transposase genes. (**A** to **F**) Northern hybridization against total RNA extracted from vegetative cells (V) as well as 11-, 24-, and 34-hour time points (T) after mixing of mating types. TBE transposase transcripts are most abundant 24 hours after mixing, when DNA rearrangements occur (*22*). 10 μg of RNA was loaded per lane and hybridized to ~1-kb DNA probes (*23*) for (A) TBE1, (B) TBE3, (D) TBE1, -2, and -3 mixed together (1:1:1 ratio), (E) TBE2 transposase, and (C and F) small-subunit ribosomal RNA (rRNA); (C) is the control for (A) and (B), and (F) is the control for (D) and (E). (**G**) Reverse transcription PCR of TBE1 (lanes 17 to 26) and TBE2 (lanes 27 to 36) transposase polyadenylate-containing mRNA from vegeta- tive cells (V) and three time points (T) during conjugation (*23*). Plus and minus symbols indicate the presence or absence of reverse transcriptase (RT), respectively. Arrowheads indicate the size of relevant markers [1-kb ladder (Invitrogen, Carlsbad, CA)].
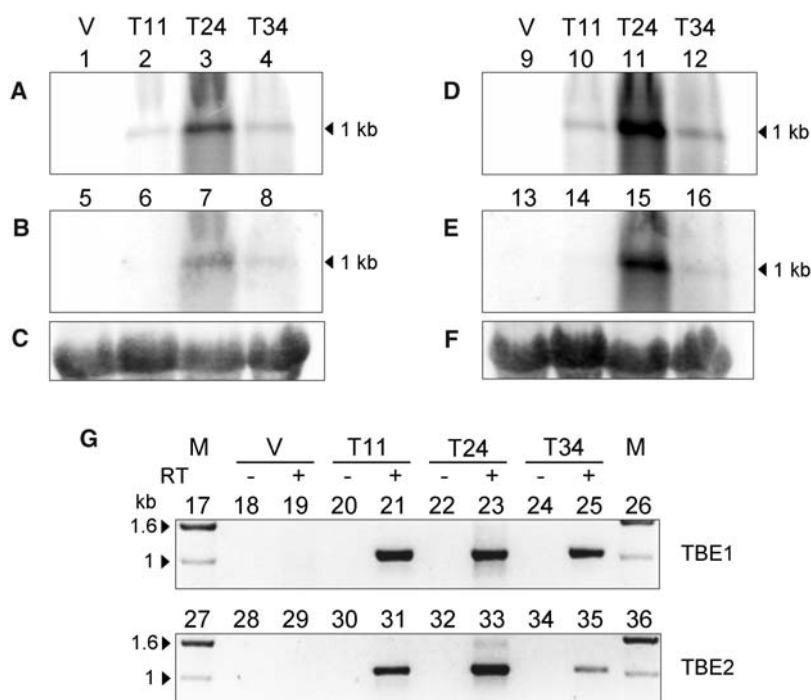
Science

Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1

Mamta Tahiliani, Kian Peng Koh, Yinghua Shen, William A. Pastor, Hozefa Bandukwala, Yevgeny Brudno, Suneet Agarwal, Lakshminarayan M. Iyer, David R. Liu, L. Aravind, and Anjana Rao

Methylation Mediation

Methylation of cytosine bases, 5-methylcytosine (5mC), in DNA plays an important regulatory role in mammalian genomes. Methylation patterns are often inherited across generations, but they can also be dynamic, suggesting that active DNA demethylation pathways exist. One such pathway, best characterized in plants, involves the removal of the 5mC base, and its replacement by C, via a DNA repair mechanism. Kriaucionis and Heintz (p. 929, published online 16 April) now show that, as well as 5mC in mammalian genomes, there are also significant amounts of 5-hydroxymethylcytosine (5hmC) in DNA of Purkinje neurons, which have large nuclei with apparently very little heterochromatin. Tahiliani *et al.* (p. 930, published online 16 April) find that the protein TET1 is capable of converting 5mC into 5hmC both in vitro and in vivo. 5-Hydroxymethylcytosine is also present in embryonic stem cells, and levels of 5hmC and TET1 show correlated variation during cell differentiation.

View the article online
https://www.science.org/doi/10.1126/science.1170116
Permissions
https://www.science.org/help/reprints-and-permissions