A TET Homologue Protein from *Coprinopsis cinerea* (CcTET) That Biochemically Converts 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine

Liang Zhang,[†,∥] Weizhong Chen,[‡,∥] Lakshminarayan M. Iyer,[§] Jennifer Hu,[†] Gloria Wang,[†] Ye Fu,[†] Miao Yu,[†] Qing Dai,[†] L. Aravind,[§] and Chuan He*[,†]

[†]Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637, United States

[‡]Department of Chemical Physics, University of Science and Technology of China, Hefei, Anhui 230026, China

[§]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, United States

Ⓢ *Supporting Information*

ABSTRACT: DNA methylation (5-methylcytosine, 5mC) plays critical biological functions in mammals and plants as a vital epigenetic marker. The Ten-Eleven translocation dioxygenases (TET1, 2, and 3) have been found to oxidize 5mC to 5-hydroxymethylcytosine (5hmC) and then to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in mammalian cells. We report herein three mushroom TET homologues from *Coprinopsis cinerea* that can mediate 5mC oxidation. Specifically, one homologue (CC1G_05589, CcTET) shows similar activity to its mammalian TET homologues. Biochemically, CcTET actively converts 5mC to 5hmC, 5fC, and 5caC under natural conditions (pH 7.0). Interestingly, CcTET also converts the majority of 5mC to 5fC under slightly acidic (pH 5.8) and neutral conditions. Kinetics analyses of the oxidation by CcTET under neutral conditions indicate that conversion of 5mC to 5hmC and 5hmC to 5fC are faster than that of 5fC to 5caC, respectively. Our results provide an example of a TET homologue in a non-mammalian organism that exhibits full 5mC-to-5caC oxidation activity and a slight preference to producing 5fC. The preferential accumulation of 5fC in the *in vitro* oxidation reactions under both neutral and acidic conditions may have biological implications for 5mC oxidation in fungi species.

D NA methylation (5mC) at the 5-position of cytosine occurs mostly on CpG dinucleotides and is widespread among eukaryotic species.[1,2] This methylation plays broad and critical biological functions especially in transcriptional regulation, gene silencing, and genome reprogramming[2−4] and exhibits dynamic distribution patterns at different developmental stages in eukaryotic genomes.[5−7] Abnormal 5mC distributions have been linked to human diseases such as cancer.[8]

Cytosine methylation is generated and maintained by a family of methyltransferases (DNMTs).[1,8] However, demethylation pathways vary among different species and are poorly understood. For instance, plants employ the Demeter family of 5mC glycosylases to excise the methylated cytosine base and replace it with an unmethylated base through the DNA base-excision repair pathway (BER).[9−14] Lacking in 5mC glycosylases, mammals must make use of different pathways. In 2009, a family of Ten-Eleven translocation dioxygenases (TET1, 2, and 3) was discovered, which led to a new paradigm of active DNA demethylation in mammals.[15] These proteins belong to the iron(II)/α-ketoglutarate (α-KG)-dependent dioxygenase family and contain a highly conserved C-terminal DNA-binding and catalytic domain (CD domain) that shares a similar, core catalytic domain with AlkB family proteins and HIF prolyl-hydroxylases.[15−20] TET proteins utilize oxygen, iron(II), and α-KG in order to oxidize the methyl group of 5mC stepwise to 5-hydroxymethylcytosine (5hmC),[15,19] 5-formylcytosine (5fC), and then 5-carboxylcytosine (5caC).[21−24] The oxidation products of 5fC and 5caC are subsequently recognized and excised by mammalian thymine DNA glycosylase (TDG) and repaired back to unmethylated cytosine through the BER pathway (Figure 1).[25−27] The 5hmC modification is relatively abundant in embryonic stem cells and neuronal cells,[28−31]
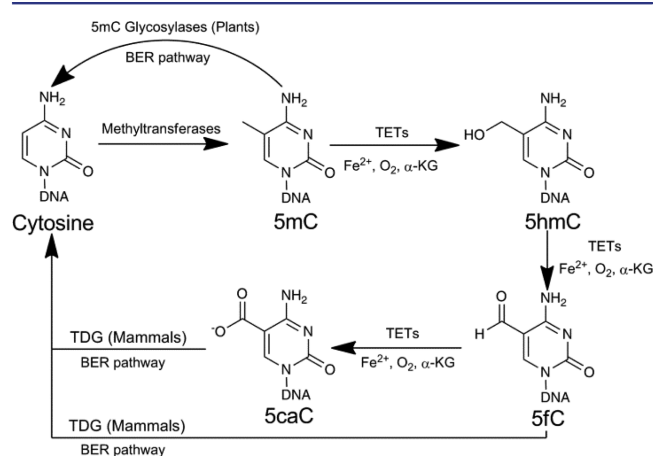

Figure 1. Mammalian cytosine methylation and the active demethylation pathway.

whereas cellular levels of 5fC and 5caC are much lower, presumably due to their constant removal through BER.[9,21,25,27]

Other than plants, mammals and certain unicellular eukaryotes, fungi contain 1−4% 5mC of the total cytosines in the genome,[32,33] but the roles of methylation in fungi remain incompletely understood. Through use of sequence profile searches, multiple TET/JBP homologues have been identified in several eukaryotes other than animals, including fungi, certain chlorophyte algae, stramenopile algae, the hetero-lobosean amoeboflagellate *Naegleria*.[35,36] One of the eight TET/JBP homologues from *Naegleria* has been recently crystallized.[34] In fungi the TET/JBP proteins are associated with a novel class of DNA transposons.[35] Here we report a newly discovered TET enzyme from *Coprinopsis cinerea* (CC1G_05589, hereafter named and referred to as CcTET), which has full activity to convert 5mC to 5hmC, 5fC, and 5caC.

C. cinerea belongs to the mushroom family *Psathyrellaceae* and is considered an important model organism of multicellular fungi. Its genome sequence was published in 2010.[36] From a total of 47 TET homologues in strain *Okayama* 7#130, we selected and expressed 12 in *E. coli* based on computational predictions.[37,38] The his-tagged recombinant proteins were purified via Ni-NTA, MonoS, and Superdex75 Gel-filtration. We employed a 5mC-oxidation assay *in vitro* to monitor the activity (see Supporting Information). A 9/11-mer duplex DNA containing a 5mCpG site on the 9mer was employed as the substrate (sequences: 5′-GAC(5mC)GGAGT-3′, 5′-GACTCCGGTCT-3′). The oxidation reactions were assayed at 37 and 25 °C under pH 5.8, 7.0, and 8.0, respectively, and the oxidation products were subsequently digested and monitored by quantitative mass spectrometry (HPLC-MS/MS). Although multiple sequence alignment demonstrated high similarity, particularly within the catalytic domain, the activity assay revealed three candidate proteins: CC1G_05589, Genbank id: XP_001838108.2; CC1G_10221, XP_001835294.1; and CC1G_02065, XP_001834329.1, which possess 5mC oxidation activity (Figure S1). All other proteins were well-expressed and could be purified; the possibility remains that some of these other proteins could be active *in vivo*.

Among the three active proteins, CC1G_05589 exhibited the highest activity; the other two moderately convert 5mC to 5hmC and 5fC with almost no 5caC observed (Figure S2). We refer to CC1G_05589 as "CcTET" here and further characterized its enzymatic activity. CcTET has 430 amino acids and shares 10.7% identity (46/430 residues) and 33.7% similarity (145/430 residues) with human TET1 (Figures S3 and S4). Moreover, the catalytic domain of CcTET is highly similar to that of mammalian (human and mouse) TET/JBP family proteins with all the catalytic residues for iron- and α-KG-binding sites conserved (such as HxD···H···R; Figure S4). Unlike its mammalian TET homologues, CcTET exhibits weak activity under reported reaction conditions (pH 8.0).[15,19,23,26,39] With an excess amount of protein (DNA:protein molar ratio of 1:2), treatment of CcTET with an 5mC-containing dsDNA produced only ~11% 5hmC and ~1.3% 5fC (Figure 2). Instead, CcTET exhibited much higher activity under neutral (pH 7.0) and slightly acidic conditions (pH 5.8) at 25 and 37 °C (Figures 2, S4, and S5). Under these conditions, CcTET generated ~28% or ~32% 5hmC and almost 2-fold more 5fC (~49% or 54%) beginning with 5mC-containing dsDNA at 37 °C, respectively; ~19% of 5caC can still be generated at neutral pH. Under slightly acidic conditions (pH 5.8), however, the
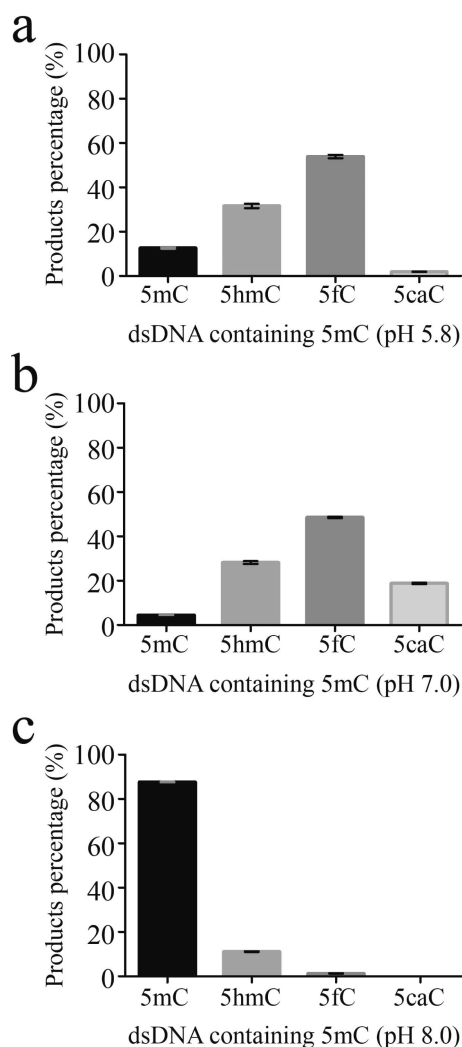

Figure 2. CcTET oxidation assays at 37 °C under pH 5.8, 7.0, and 8.0. The reaction mix contains 1 μM 5mC containing dsDNA, 2 μM CcTET, 50 mM Bis-Tris, pH 5.8 (or HEPES, pH 7.0 or 8.0), 75 μM Fe(NH$_4$)$_2$(SO$_4$)$_2$, 2 mM ascorbic acid, 1 mM α-KG, and 2.5 mM DTT for 30 min at 37 °C and then quenched by heating to 100 °C for 5 min and immediately cooled in an ice bath.

reaction almost stalls at the 5fC stage, generating only ~1.9% 5caC. Additionally, CcTET could also oxidize 5mC in single-stranded DNA (ssDNA) with similar oxidation activity and product contents as dsDNA (Figure S6).[40] In contrast, the catalytic domain of mouse TET1 (mTET1) also exhibits higher activity at pH 5.8 and 7.0 than at pH 8.0 at 37 °C. mTET1 converts almost all 5mC to 5caC under slightly acidic and neutral conditions as well as under pH 8.0 with ATP *in vitro* (Figure S7).[22,39] These biochemical data indicate that CcTET has full activity to convert 5mC to 5hmC, 5fC, and 5caC on both of single stranded and duplex DNA under neutral condition and is mostly inactive at pH 8.0 even with the presence of ATP *in vitro* (Figure S8). Notably, the accumulation of 5fC production by CcTET observed under both neutral and acidic reaction conditions may have interesting biological implications as 5fC accumulates in certain mammalian cells *in vivo* and could exhibit important functions.[41,42] Interestingly, CcTET also exhibits weak thymine oxidation activity under pH 7.0 on A·T base pair in duplex DNA substrate and oxidizes thymine to 5-hydroxymethylur-

idine (5hmU) in our activity assay (Figure S9). This activity toward thymine is much lower than oxidation of 5mC. However, this is consistent with the recent prediction regarding existence of hypermodified thymines in fungi such as Coprinopsis38 and raises the possibility that ccTET, despite being in a distinct clade from JBP, has more JBP-like activity that mammalian TET.

To further probe the intriguing 5mC oxidation activity, we also performed oxidation assays of CcTET on genomic DNA isolated from mouse embryonic stem cells (mESC) under neutral and slightly acidic conditions with similar results observed (Figures S10−S12). We then performed electrophoretic mobility shift assays (EMSA) against cytosine derivatives (C, 5mC, 5hmC, 5fC, and 5caC) contained in a 22mer dsDNA sequence (Figure S13). The results show that CcTET could bind to all these cytosine derivatives with very similar binding affinity (~125 nM), suggesting that CcTET does not have noticeable selective preference to these modifications. It should be noted that the mammalian TET enzymes oxidize 5hmC all the way to 5caC in vitro with minimum 5fC observed (Figure S7).[39] In the case of CcTET it is possible that 5fC bound in the active site of the enzyme cannot be effectively hydrolyzed to form hemiacetal, which could be critical for continued oxidation to 5caC.[43]

To further understand the catalytic mechanism of CcTET, we performed kinetic analyses on the reactions of 5mC to 5hmC, 5hmC to 5fC, and 5fC to 5caC at 37 °C using dsDNA under neutral conditions (pH 7.0), employing quantitative mass spectrometry. We first collected the reaction progress curve of substrate and fraction products (5hmC, 5fC, and 5caC) versus incubation times at 37 °C (Figure S14). Like mTET1,[22] the CcTET-catalyzed reaction plateaus after ~10 min of incubation; inactivation of CcTET at the reaction temperature (37 °C) is less likely the reason (Figure S15). The progress curve appears linear within 2.5 min, allowing for initial rates to be derived (Figure S16). We then performed kinetic study within 2.5 min at 37 °C and quenched the reactions at time points of 0, 30, 60, 90, 120, and 150 s. As shown in Figure 3 and Table S1, the oxidation activity of 5mC to 5hmC, 5hmC to 5fC, and 5fC to 5caC under pH 7.0 at 37 °C gradually decreases. Under these experimental conditions, the maximal rate of the oxidation of 5mC to 5hmC is slightly faster than that of 5hmC to 5fC, which is faster than that of the conversion of 5fC to 5caC. The half-maximal rate values ($1/2\ V_{max}$) of 5mC to 5hmC, 5hmC to 5fC, and 5fC to 5caC were obtained at similar substrate concentrations, suggesting that CcTET lacks binding preference to these modified cytosine bases, which is consistent with the EMSA results (Figure S13). The progress curves of 5mC to 5hmC and 5hmC to 5fC could be influenced by the subsequent oxidation step. However, the subsequent oxidation steps show slower reaction rates which may not significantly affect the kinetic analysis of the previous step (Figure S16).

The discovery of CcTET and the other two homologues as well as the characterization may have several implications: (1) It suggests that fungal TET homologues can catalyze the oxidation of 5mC to further modified derivatives; these conversions might have significant biological functions as epigenetic marks or as intermediates in active demethylation in these fungi; (2) the fact that CcTET yields more 5fC over 5caC in vitro might suggest that different TET enzymes have distinctly tuned their active sites in order to generate different reaction products; this activity tuning perhaps arises from 5fC
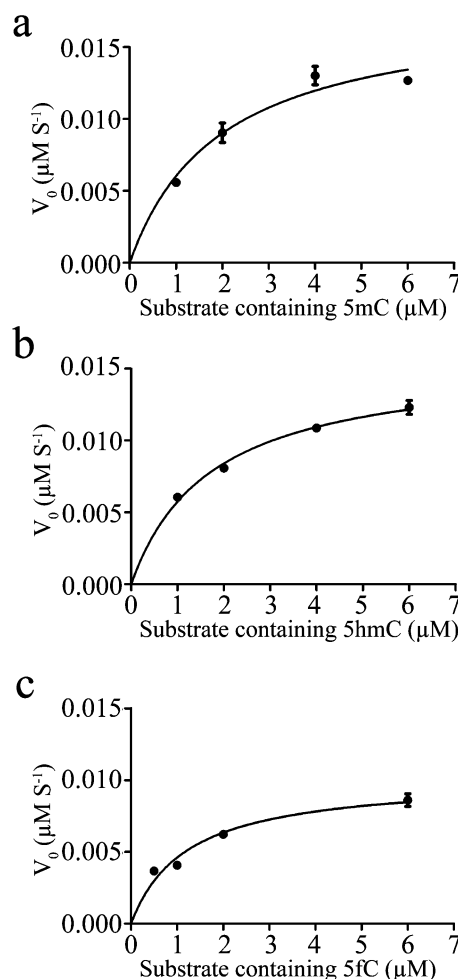

Figure 3. Michaelis−Menten's plot of the kinetics of reactions of 5mC to 5hmC, 5hmC to 5fC, and 5fC to 5caC catalyzed by CcTET at 37 °C in pH 7.0 solution.

or its derivatives in genomic DNA playing specific roles in these fungi; and (3) mouse TET proteins have been shown to function as excellent tools to convert 5mC to either 5hmC (through in situ trapping) or 5caC in sequencing applications.[39,44,45] The production of 5fC by CcTET under both neutral and acidic reaction conditions could be explored in order to develop new sequencing strategies for cytosine modifications. In addition, the molecular mechanism of CcTET-mediated oxidation can be compared with that of the mammalian mechanism in the future when structures of these enzymes become available. Importantly, these findings offer the first biochemical glimpses of the TET homologues, which display major lineage-specific expansions in several fungi and appear to be notable contributors to the genomic and epigenetic landscape of these organisms.[35]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Experimental details and kinetic profiles. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

chuanhe@uchicago.edu

Author Contributions

‖These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Law, J. A.; Jacobsen, S. E. *Nat. Rev. Genet.* **2010**, *11*, 204.
(2) Wu, S. C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607.
(3) Bhutani, N.; Burns, D. M.; Blau, H. M. *Cell* **2011**, *146*, 866.
(4) Klose, R. J.; Bird, A. P. *Trends Biochem. Sci.* **2006**, *31*, 89.
(5) Weber, M.; Schubeler, D. *Curr. Opin. Cell Biol.* **2007**, *19*, 273.
(6) Reik, W.; Dean, W.; Walter, J. *Science* **2001**, *293*, 1089.
(7) Sasaki, H.; Matsui, Y. *Nat. Rev. Genet.* **2008**, *9*, 129.
(8) Robertson, K. D.; Wolffe, A. P. *Nat. Rev. Genet.* **2000**, *1*, 11.
(9) Feng, S.; Jacobsen, S. E.; Reik, W. *Science* **2010**, *330*, 622.
(10) Suzuki, M. M.; Bird, A. *Nat. Rev. Genet.* **2008**, *9*, 465.
(11) Gehring, M.; Huh, J. H.; Hsieh, T. F.; Penterman, J.; Choi, Y.; Harada, J. J.; Goldberg, R. B.; Fischer, R. L. *Cell* **2006**, *124*, 495.
(12) Choi, Y.; Gehring, M.; Johnson, L.; Hannon, M.; Harada, J. J.; Goldberg, R. B.; Jacobsen, S. E.; Fischer, R. L. *Cell* **2002**, *110*, 33.
(13) Hsieh, T. F.; Ibarra, C. A.; Silva, P.; Zemach, A.; Eshed-Williams, L.; Fischer, R. L.; Zilberman, D. *Science* **2009**, *324*, 1451.
(14) Mok, Y. G.; Uzawa, R.; Lee, J.; Weiner, G. M.; Eichman, B. F.; Fischer, R. L.; Huh, J. H. *Proc. Natl. Acad Sci. U.S.A.* **2010**, *107*, 19225.
(15) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
(16) Yi, C.; Yang, C. G.; He, C. *Acc. Chem. Res.* **2009**, *42*, 519.
(17) Hausinger, R. P. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 21.
(18) Sedgwick, B.; Bates, P. A.; Paik, J.; Jacobs, S. C.; Lindahl, T. *DNA Repair* **2007**, *6*, 429.
(19) Ito, S.; D'Alessio, A. C.; Taranova, O. V.; Hong, K.; Sowers, L. C.; Zhang, Y. *Nature* **2010**, *466*, 1129.
(20) Elkins, J. M.; Hewitson, K. S.; McNeill, L. A.; Seibel, J. F.; Schlemminger, I.; Pugh, C. W.; Ratcliffe, P. J.; Schofield, C. J. *J. Biol. Chem.* **2003**, *278*, 1802.
(21) Pfaffeneder, T.; Hackner, B.; Truss, M.; Munzel, M.; Muller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. *Angew. Chem., Int. Ed. Engl.* **2011**, *50*, 7008.
(22) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300.
(23) Krumova, P.; Weishaupt, J. H. *Cell. Mol. Life Sci.* **2013**, *70*, 2123.
(24) Yin, R.; Mao, S. Q.; Zhao, B.; Chong, Z.; Yang, Y.; Zhao, C.; Zhang, D.; Huang, H.; Gao, J.; Li, Z.; Jiao, Y.; Li, C.; Liu, S.; Wu, D.; Gu, W.; Yang, Y. G.; Xu, G. L.; Wang, H. *J. Am. Chem. Soc.* **2013**, *135*, 10396.
(25) He, Y. F.; Li, B. Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C. X.; Zhang, K.; He, C.; Xu, G. L. *Science* **2011**, *333*, 1303.
(26) Zhang, L.; Lu, X.; Lu, J.; Liang, H.; Dai, Q.; Xu, G. L.; Luo, C.; Jiang, H.; He, C. *Nat. Chem. Biol.* **2012**, *8*, 328.
(27) Maiti, A.; Drohat, A. C. *J. Biol. Chem.* **2011**, *286*, 35334.
(28) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929.
(29) Pastor, W. A.; Pape, U. J.; Huang, Y.; Henderson, H. R.; Lister, R.; Ko, M.; McLoughlin, E. M.; Brudno, Y.; Mahapatra, S.; Kapranov, P.; Tahiliani, M.; Daley, G. Q.; Liu, X. S.; Ecker, J. R.; Milos, P. M.; Agarwal, S.; Rao, A. *Nature* **2011**, *473*, 394.
(30) Globisch, D.; Munzel, M.; Muller, M.; Michalakis, S.; Wagner, M.; Koch, S.; Bruckl, T.; Biel, M.; Carell, T. *PLoS ONE* **2010**, *5*, e15367.

(31) Mellen, M.; Ayata, P.; Dewell, S.; Kriaucionis, S.; Heintz, N. *Cell* **2012**, *151*, 1417.
(32) Binz, T. *Mycologia* **1998**, *90*, 785.
(33) Zemach, A.; McDaniel, I. E.; Silva, P.; Zilberman, D. *Science* **2010**, *328*, 916.
(34) Hashimoto, H.; Pais, J. E.; Zhang, X.; Saleh, L.; Fu, Z. Q.; Dai, N.; Correa, I. R.; Zheng, Y.; Cheng, X. *Nature* **2013**, *506*, 391.
(35) Iyer, L. M.; Zhang, D.; de Souza, R. F.; Pukkila, P. J.; Rao, A.; Aravind, L. *Proc. Natl. Acad Sci. U.S.A.* **2014**, *111*, 1676.
(36) Stajich, J. E.; Wilke, S. K.; Ahren, D.; Au, C. H.; Birren, B. W.; Borodovsky, M.; Burns, C.; Canback, B.; Casselton, L. A.; Cheng, C. K.; Deng, J.; Dietrich, F. S.; Fargo, D. C.; Farman, M. L.; Gathman, A. C.; Goldberg, J.; Guigo, R.; Hoegger, P. J.; Hooker, J. B.; Huggins, A.; James, T. Y.; Kamada, T.; Kilaru, S.; Kodira, C.; Kues, U.; Kupfer, D.; Kwan, H. S.; Lomsadze, A.; Li, W.; Lilly, W. W.; Ma, L. J.; Mackey, A. J.; Manning, G.; Martin, F.; Muraguchi, H.; Natvig, D. O.; Palmerini, H.; Ramesh, M. A.; Rehmeyer, C. J.; Roe, B. A.; Shenoy, N.; Stanke, M.; Ter-Hovhannisyan, V.; Tunlid, A.; Velagapudi, R.; Vision, T. J.; Zeng, Q.; Zolan, M. E.; Pukkila, P. J. *Proc. Natl. Acad Sci. U.S.A.* **2010**, *107*, 11889.
(37) Iyer, L. M.; Tahiliani, M.; Rao, A.; Aravind, L. *Cell Cycle* **2009**, *8*, 1698.
(38) Iyer, L. M.; Zhang, D.; Burroughs, A. M.; Aravind, L. *Nucleic Acids Res.* **2013**, *41*, 7635.
(39) Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C. X.; Zhang, L.; Kim, A.; Li, X.; Dai, Q.; Shen, Y.; Park, B.; Min, J. H.; Jin, P.; Ren, B.; He, C. *Cell* **2012**, *149*, 1368.
(40) Zhang, L.; Yu, M.; He, C. *Acta Chim. Sinica* **2012**, *70*, 2123.
(41) Song, C. X.; Szulwach, K. E.; Dai, Q.; Fu, Y.; Mao, S. Q.; Lin, L.; Street, C.; Li, Y.; Poidevin, M.; Wu, H.; Gao, J.; Liu, P.; Li, L.; Xu, G. L.; Jin, P.; He, C. *Cell* **2013**, *153*, 678.
(42) Shen, L.; Wu, H.; Diep, D.; Yamaguchi, S.; D'Alessio, A. C.; Fung, H. L.; Zhang, K.; Zhang, Y. *Cell* **2013**, *153*, 692.
(43) Fu, Y.; He, C. *Curr. Opin. Chem. Biol.* **2012**, *16*, 516.
(44) Zhang, L.; Szulwach, K. E.; Hon, G. C.; Song, C. X.; Park, B.; Yu, M.; Lu, X.; Dai, Q.; Wang, X.; Street, C. R.; Tan, H.; Min, J. H.; Ren, B.; Jin, P.; He, C. *Nat. Comm.* **2013**, *4*, 1517.
(45) Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C. X.; Jin, P.; Ren, B.; He, C. *Nat. Protoc.* **2012**, *7*, 2159.