nature neuroscience

Role of Tet1 and 5-hydroxymethylcytosine in cocaine action

Jian Feng[1], Ningyi Shao[1,8], Keith E Szulwach[2,8], Vincent Vialou[1,3,8], Jimmy Huynh[1], Chun Zhong[4], Thuc Le[5], Deveroux Ferguson[1], Michael E Cahill[1], Yujing Li[2], Ja Wook Koo[1], Efrain Ribeiro[1], Benoit Labonte[1], Benjamin M Laitman[1], David Estey[1], Victoria Stockman[1], Pamela Kennedy[1], Thomas Couroussé[3], Isaac Mensah[1], Gustavo Turecki[6], Kym F Faull[7], Guo-li Ming[4], Hongjun Song[4], Guoping Fan[5], Patrizia Casaccia[1], Li Shen[1], Peng Jin[2] & Eric J Nestler[1]

Ten-eleven translocation (TET) enzymes mediate the conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), which is enriched in brain, and its ultimate DNA demethylation. However, the influence of TET and 5hmC on gene transcription in brain remains elusive. We found that ten-eleven translocation protein 1 (TET1) was downregulated in mouse nucleus accumbens (NAc), a key brain reward structure, by repeated cocaine administration, which enhanced behavioral responses to cocaine. We then identified 5hmC induction in putative enhancers and coding regions of genes that have pivotal roles in drug addiction. Such induction of 5hmC, which occurred similarly following TET1 knockdown alone, correlated with increased expression of these genes as well as with their alternative splicing in response to cocaine administration. In addition, 5hmC alterations at certain loci persisted for at least 1 month after cocaine exposure. Together, these reveal a previously unknown epigenetic mechanism of cocaine action and provide new insight into how 5hmC regulates transcription in brain *in vivo*.

© 2015 Nature America, Inc. All rights reserved.

npg

DNA methylation is an epigenetic mechanism by which methyl groups are covalently coupled to the C-5 position of cytosine (5mC) predominantly at CpG dinucleotides[1]. It is generally accepted that DNA methylation is relatively stable and mediates long-term gene silencing. This presents an appealing mechanism for long-lasting transcriptional regulation underlying neural plasticity associated with learning and memory and neuropsychiatric disorders. However, accumulating evidence indicates that DNA methylation in brain is reversible[2–5], which suggests the existence of DNA demethylation machinery. Although several candidates have been proposed as DNA demethylases[6] in mammals, a major breakthrough came in 2009, when TET1 was recognized to convert 5mC to 5hmC (refs. 7,8). Soon after, two other TET family members, TET2 and TET3, were also shown to have 5mC hydroxylase activity[9,10]. It was further revealed that all three TETs successively oxidize 5mC to 5hmC, 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[11,12]. Thus, 5mC oxidation is a plausible DNA demethylation mechanism. In fact, studies have discovered several pathways leading to 5hmC-mediated DNA demethylation[13–15]. For example, TET-catalyzed 5hmC conversion into 5fC and 5caC can be efficiently removed from DNA by thymine-DNA glycosylase. Subsequent repair of the resulting abasic site via base excision repair can generate an unmethylated cytosine. Alternatively, 5hmC deamination can remove 5hmC and replace it with unmethylated cytosine[16].

We have just started to understand the distribution and function of these forms of DNA epigenetic modifications and TET oxidases in the genome, which are essential in a range of biological processes such as embryonic development, stem cell function and cancer formation[13–15]. Of note, 5hmC is most abundant in brain compared with other organs[7,17,18], suggesting that is important in neural function[19]. Indeed, emerging evidence indicates that TET1 regulates active hippocampal DNA demethylation[16,20,21], cognition[22], and learning and memory[20,21,23]. However, the influence of TET proteins and 5hmC on the regulation of gene transcription in brain and their role in brain disorders remain largely unknown.

We set out to determine whether TET proteins and 5hmC are involved in the epigenetic regulation of cocaine action. Repeated cocaine exposure induced persistent changes in gene expression in the nucleus accumbens (NAc)[24], a key brain structure of the reward circuitry. Epigenetic mechanisms are important for this process as well as in downstream neural and behavioral plasticity[25,26]. We found decreased expression of TET1, but not TET2 or TET3, in NAc after repeated cocaine administration. Using viral manipulations, we established that decreased TET1 served to enhance behavioral responses to the drug. In concert with TET1 downregulation, repeated cocaine increased the enrichment of 5hmC at a large subset of genes that are involved in drug addiction. Such changes were concentrated at both putative enhancers

[1]Fishberg Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [2]Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, USA. [3]Institut National de la Santé et de la Recherhe Médicale (INSERM) U1130, CNRS UMR8246, UPMC UM18, Neuroscience Paris Seine, Paris, France. [4]Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [5]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California, USA. [6]The McGill Group for Suicide Studies, Douglas Hospital Research Centre, McGill University, Montreal, Canada. [7]Pasarow Mass Spectrometry Laboratory, Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, California, USA. [8]These authors contributed equally to this work.
Correspondence should be addressed to E.J.N. (eric.nestler@mssm.edu).

Received 6 January; accepted 17 February; published online 16 March 2015; doi:10.1038/nn.3976

and gene bodies, and correlated with increased expression of these genes as well as with their alternative splicing. Notably, we found that cocaine-induced increases in 5hmC at representative loci in NAc were mimicked following *Tet1* knockdown alone. In sum, our findings not only advance our understanding of the epigenetic mechanisms involved in cocaine action, but also provide fundamentally new insight into how 5hmC regulates gene expression in the brain.

RESULTS

Repeated cocaine decreases TET1 in NAc to enhance behavioral responses

We first examined *Tet* mRNA regulation in mouse NAc 24 h after repeated cocaine exposure. Quantitative PCR (qPCR) revealed that, of the three known *Tet* family members, only *Tet1* was downregulated (**Fig. 1a**). Quantitative western blotting and immunohistochemistry revealed a concomitant decrease in TET1 protein, but not TET2 or TET3, in NAc after repeated cocaine (**Fig. 1b**,**c** and **Supplementary Fig. 1a**,**b**). Notably, we found a ~40% decrease in *TET1* mRNA in the NAc of cocaine-addicted humans examined postmortem (**Fig. 1d**). These data demonstrate that TET1 expression is subject to dynamic regulation in the adult brain and implicate TET1 in cocaine action.

To investigate directly whether altered levels of TET1 in NAc regulate behavioral responses to cocaine, we used stereotaxic viral manipulations to express *Tet1* shRNA and knockdown TET1 selectively in the adult NAc (**Supplementary Figs. 1c** and **2a**,**b**). We then analyzed the mice in an unbiased cocaine conditioned place preference procedure, which provides an indirect measure of drug reward. *Tet1* knockdown with either of two independent shRNA constructs robustly enhanced cocaine place conditioning (**Fig. 1e**,**f**). Conversely, we used a previously validated AAV-TET1 vector to express TET1 (ref. 16) in NAc (**Supplementary Fig. 2c**) and found decreased cocaine preference (**Fig. 1g**), which indicates that TET1 expression in the NAc is sufficient to suppress cocaine reward. Together, these data suggest that TET1 normally functions to negatively regulate cocaine reward and that cocaine-induced suppression of TET1 in NAc contributes to enhanced drug sensitivity.

5hmC profiling in NAc after repeated cocaine administration

Given that TET1 oxidizes 5mC into 5hmC, we next investigated cocaine regulation of these DNA modifications in NAc. First, we quantified total levels of 5mC and 5hmC[5,27] and found no global change in either (**Fig. 2a**). This suggests that any alterations in 5hmC may be locus specific. Alternatively, given the ~2-fold greater abundance of *Tet2* and *Tet3* mRNAs in NAc (**Supplementary Fig. 3**), it is conceivable that they compensate for TET1 downregulation despite their lack of regulation by cocaine (**Fig. 1a**,**c**). To test these possibilities, we mapped 5hmC genome-wide using a 5hmC chemical labeling capture protocol[19] that provided high-quality sequencing results (**Supplementary Table 1**).

In total, we identified 208,801 5hmC peak regions from NAc of saline-treated control mice and 226,185 peaks from cocaine-treated mice. Although the gross chromosomal distribution of 5hmC peak regions was equivalent between saline- and cocaine-treated conditions, we found that sex chromosomes had extremely low 5hmC enrichment (**Supplementary Fig. 4a**). We also identified 11,511 regions that displayed differential levels of 5hmC after repeated cocaine (**Supplementary Table 2**), which appeared to be evenly distributed across all autosomes (**Supplementary Fig. 4b**). These 5hmC differential regions were heavily enriched in gene bodies (~55%) and intergenic regions (~34%), as determined by calculating the percentage of differential region counts across all genomic features (**Fig. 2b** and **Supplementary Fig. 4c**). In addition, we measured the density of these 5hmC alterations across various genomic regions and found intergenic regions, gene promoters and gene bodies among the top categories (**Fig. 2c**). Thus, we speculated that 5hmC might have functional roles in both intergenic and genic regions in cocaine action.

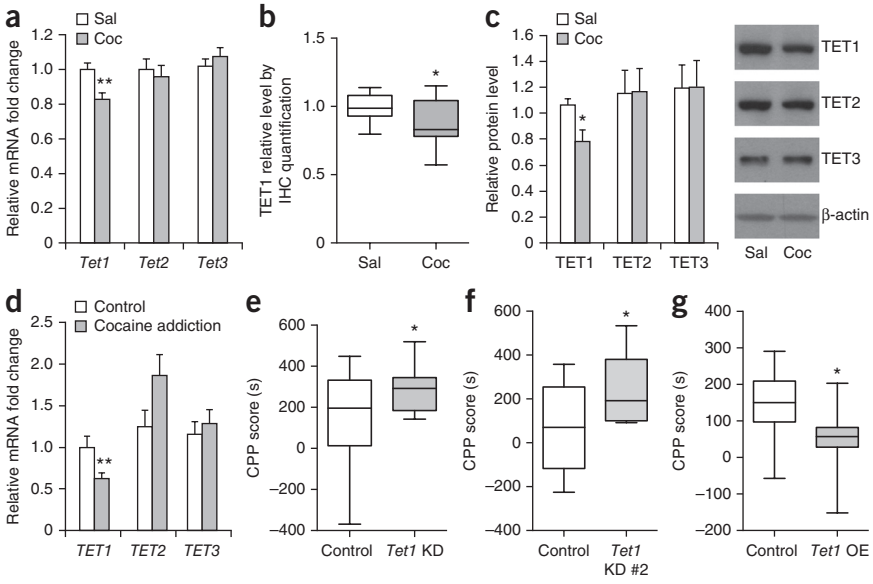

Figure 1 TET1 expression is decreased in NAc after repeated cocaine and negatively regulates behavioral responses to cocaine. (**a**) qPCR analysis showing relative transcription of *Tet1*, *Tet2* and *Tet3* mRNA in NAc 24 h after repeated cocaine (unpaired *t* test, *Tet1*: $P = 0.002$, $t(34) = 3.321$, $n = 18$ per group; *Tet2*: $P = 0.186$, $t(33) = 1.350$, saline $n = 18$, cocaine $n = 17$; *Tet3*: $P = 0.393$, $t(46) = 0.861$, $n = 24$ per group). (**b**) Quantification of TET1 in NAc by immunohistochemistry 24 h after repeated cocaine (unpaired *t* test, $P = 0.035$, $t(22) = 2.245$, $n = 12$ per group). (**c**) Western blot analysis revealed a selective decrease in TET1 protein 24 h after repeated cocaine (unpaired *t* test, TET1: Welch correction, $P = 0.019$, $t(25) = 2.500$, $n = 18$ per group; TET2: $P = 0.970$, $t(28) = 0.038$, $n = 16$ per group; TET3: $P = 0.972$, $t(35) = 0.035$, saline $n = 19$, cocaine $n = 18$). A representative blot is shown with β-actin as a loading control. (**d**) qPCR analysis showed a selective decrease in *TET1* mRNA, but not *TET2* or *TET3*, in NAc of humans addicted to cocaine (unpaired *t* test, *TET1*: $P = 0.006$, $t(38) = 2.913$, control $n = 21$, addicted $n = 19$; *TET2*: $P = 0.065$, $t(36) = 1.903$, $n = 19$ per group; *TET3*: $P = 0.463$, $t(34) = 0.743$, $n = 18$ per group.). (**e**) Conditioned place preference (CPP) for cocaine with viral *Tet1* shRNA knockdown (KD) construct#1 in NAc (unpaired *t* test Welch correction, $P = 0.038$, $t(21) = 2.221$, $n = 15$ for control and 12 for KD). (**f**) CPP for cocaine with viral *Tet1* shRNA KD construct#2 in NAc (unpaired *t* test, $P = 0.031$, $t(22) = 2.303$, $n = 12$ per group). (**g**) CPP for cocaine with AAV-TET1 overexpression (OE) in NAc (unpaired *t* test, $P = 0.027$, $t(23) = 2.358$, $n = 14$ for control and 11 for OE group). Data are presented as mean ± s.e.m. in **a**, **c** and **d**. Box plots present, in ascending order, minimum sample value, first quartile, median, third quartile, maximum sample value. $^*P < 0.05$, $^{**}P < 0.01$. Sal, saline; Coc, cocaine.

© 2015 Nature America, Inc. All rights reserved.

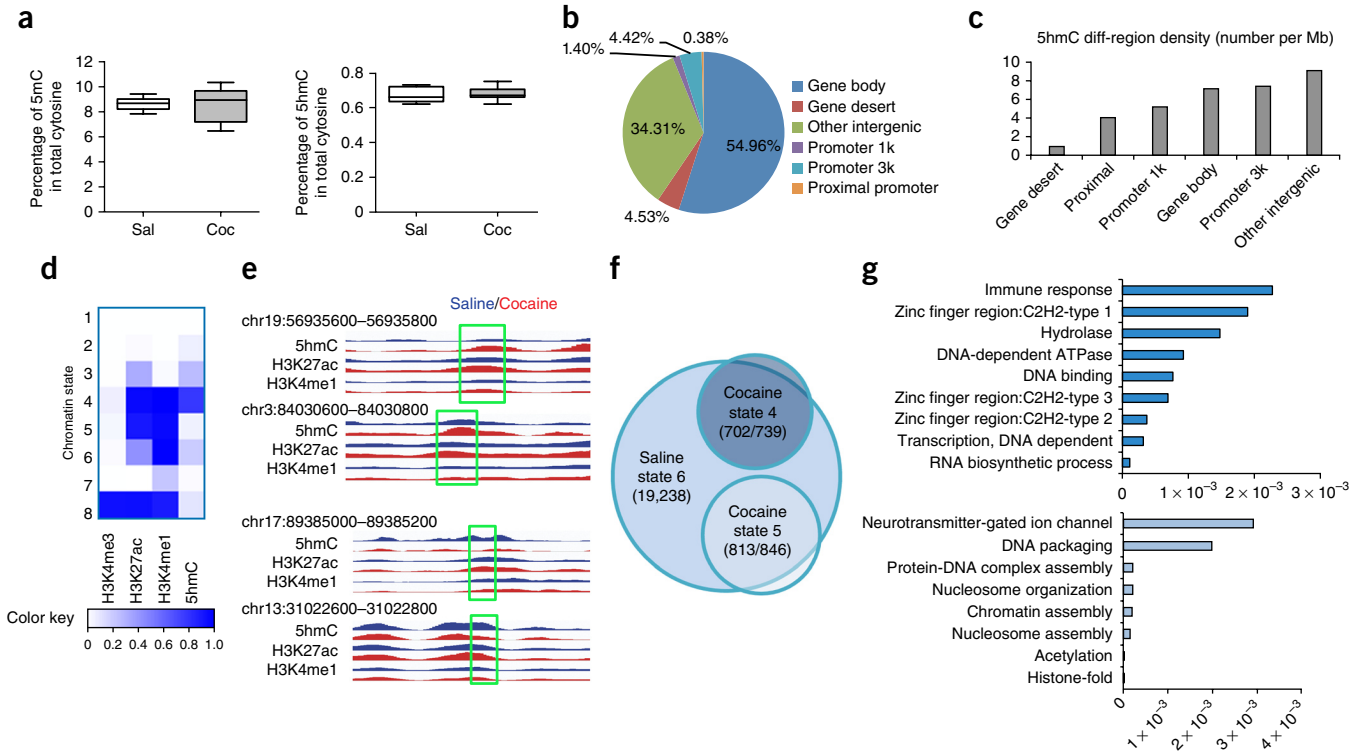

Figure 2 Repeated cocaine induces 5hmC alterations in NAc. (**a**) Quantitative analysis of 5mC and 5hmC using LC-ESI-MS per MS (unpaired *t* test, 5mC: $P = 0.990$, $t(16) = 0.018$; 5hmC: $P = 0.787$, $t(16) = 0.295$; $n = 9$ per group). Box plots are presented as in **Figure 1**. (**b**) Genomic distribution of sites that show cocaine-induced changes in 5hmC enrichment. (**c**) 5hmC differential region density is presented as the number of cocaine-induced differential regions per Mb sequence. (**d**) Chromatin state is generated to define combinatorial patterns of H3K4me3, H3K27ac, H3K4me1 and 5hmC. Eight distinct states are shown. Color key reflects the emission probabilities of a hidden Markov model, which denotes the frequencies of chromatin mark presences. (**e**) Representative tracks in saline (blue) or cocaine (red) illustrate a chromatin state switch from state 6 in saline to state 4 in cocaine (top two) or to state 5 in cocaine (bottom two). The regions that display enrichment changes in 5hmC and H3K27ac are highlighted with a box and denoted with the corresponding genomic coordinates. *y* axis reflects normalized read counts that are set to the same scale for saline and cocaine. (**f**) The Venn diagram demonstrates that the vast majority of cocaine-specific putative enhancer genes (702 of 739 for state 4 and 813 of 846 for state 5) had a chromatin state switch at the putative enhancer site from state 6 to state 4 or 5 after repeated cocaine. (**g**) List of GO terms that were most significantly enriched in cocaine-specific state 4 (top) and cocaine-specific state 5 putative enhancer genes (bottom); *x* axes show *P* values.

© 2015 Nature America, Inc. All rights reserved.

npg

5hmC dynamics at putative enhancers

We first characterized 5hmC dynamics at putative enhancer regions. Enhancers are DNA regulatory elements that are critical in gene expression, with many residing long distances away from gene promoters. It is estimated that hundreds of thousands of enhancers exist in the genome, vastly outnumbering the ~20,000 coding genes. Enhancers are also associated with characteristic histone modification patterns[28]. For example, H3K27ac and H3K4me1 co-binding has been widely recognized as a signature of active enhancers that engage in transcription regulation[28–32]. In accordance with these studies, to define putative enhancers in NAc, we performed chromatin immunoprecipitation sequencing (ChIP-seq) for H3K4me1 and H3K27ac (**Supplementary Table 1**). We also included a previously reported ChIP-seq data set for H3K4me3 (ref. 33), which is highly enriched at promoters, to exclude promoters from our predicted enhancers.

When we plotted our ChIP-seq data of H3K4me1 and H3K27ac over putative mouse brain enhancers derived from the ENCODE database (http://genome.ucsc.edu/ENCODE/), both display enriched binding at enhancer regions (**Supplementary Fig. 5**). This analysis suggests good antibody specificity and sequencing quality. Similarly, 5hmC showed enrichment at these putative enhancer regions genome-wide. We then performed a study on chromatin states[28,32], or the combination of epigenetic modifications, using ChromHMM[32]. We applied combinatorial patterns of H3K4me1 and H3K27ac, together with 5hmC, to define several distinct chromatin states (Online Methods and **Fig. 2d**)[32]. We first excluded state 8 from our putative enhancer analysis because of its high levels of H3K4me3, which reflect promoter presence (**Fig. 2d**). We then focused on the chromatin states that were characterized by enhancer marks H3K4me1 and H3K27ac, as well as 5hmC (states 4, 5 and 6; **Fig. 2d**). To further validate our definition of putative enhancers by H3K4me1 and H3K27ac ChIP-seq through ChromHMM, we applied a second algorithm (RFECS, random forest based enhancer identification from chromatin states) that reportedly accurately predicts enhancers based solely on H3K4me1 and H3K27ac ChIP-seq[28]. We found that the enhancer predictions from the two algorithms had >70% overlap, which provides validity for our enhancer predication approach. We next performed ChIP-qPCR to validate H3K4me1 and H3K27ac enrichment at multiple predicted putative enhancer sites (**Supplementary Fig. 6**). Notably, these putative enhancer sites exhibited various degrees of enrichment of P300 (**Supplementary Fig. 6**), a transcription co-activator that also marks enhancer regions. Thus, these findings further support our method of identifying putative enhancers.

Our analyses revealed robust regulation of chromatin states at putative enhancer regions in NAc by cocaine. A subgroup of putative enhancers displayed increased prevalence of H3K27ac together with either increased 5hmC (switch from state 6 to 4) or decreased 5hmC (switch from state 6 to state 5) after repeated cocaine (**Fig. 2d**,**e**).

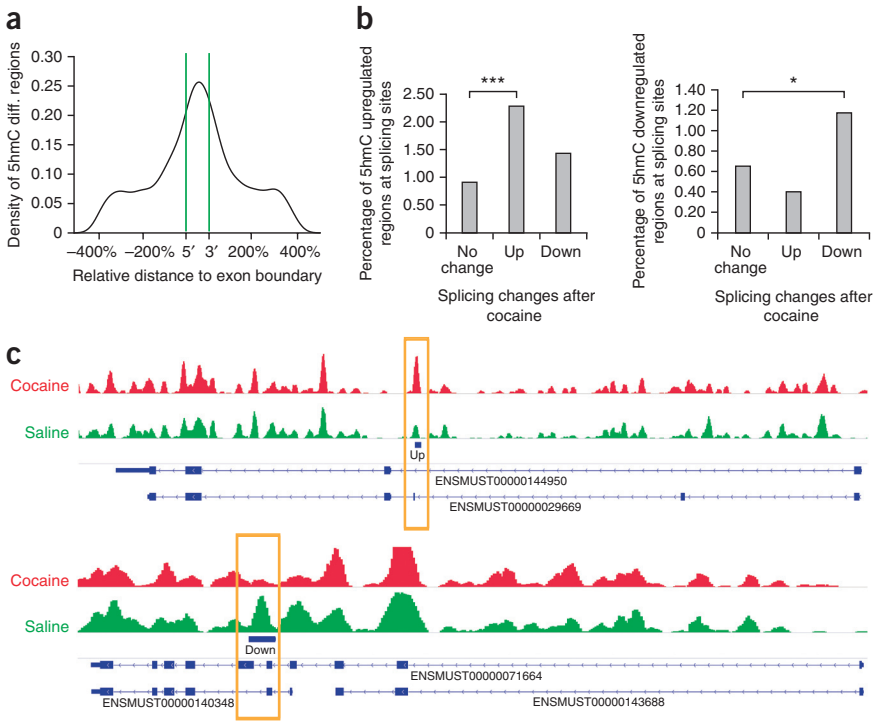

Figure 3 5hmC alterations positively correlate with alternative splicing regulation. (**a**) 5hmC differential region density plot at exon regions. The *x* axis represents an exon at the center with flanking regions that are of the size of 400% of the central exon. Green lines illustrate exon boundaries. (**b**) Percentage enrichment of upregulated 5hmC regions (left) or downregulated 5hmC regions (right) at splicing sites of increased (up), decreased (down) and not significantly changed (no change) splicing isoforms in response to cocaine (Fisher's exact test, $^{***}P = 4.36 \times 10^{-6}$, $^{*}P = 0.022$). (**c**) Representative 5hmC tracks in saline (green) and cocaine (red) conditions that correlate with altered expression of a splicing isoform. A yellow box highlights the altered 5hmC after repeated cocaine. In the upper sample, increased 5hmC was associated with upregulation of the splicing isoform (transcript ID: ENSMUST00000029669, chr3:151845628–151874378, RNAseq $\log_2$ fold change = 1.079, *Q* value = 0.0085). In the lower sample, decreased 5hmC was associated with downregulation of the splicing isoform (transcript ID: ENSMUST00000071664, chr3:94836397–94846802, RNAseq $\log_2$ fold change = –0.507, *Q* value = 0.0005).

© 2015 Nature America, Inc. All rights reserved.

Given that enhancers can regulate expression of neighboring protein-coding genes, we assigned them to the nearest genes as plausible targets (see Discussion for limitations of this approach). We recognized 739 and 846 cocaine-specific chromatin state 4 and state 5 putative enhancer genes, respectively (**Supplementary Table 3**). Notably, the vast majority (>90%) of these genes' enhancers were under state 6 at saline condition (**Fig. 2f**), which indicates robust 5hmC dynamics—a gain in 5hmC in switching from state 6 to state 4 or a loss of 5hmC in switching from state 6 to state 5—after cocaine administration. Gene ontology (GO) analysis of cocaine-specific state 4 and state 5 putative enhancer genes revealed enrichment in several meaningful categories (**Fig. 2g** and **Supplementary Table 4**). For example, cocaine-specific state 4 putative enhancer genes included C2H2 zinc finger proteins that bind to methylated DNA[34] and immune genes that are implicated in neuronal plasticity[35]. In contrast, cocaine-specific state 5 putative enhancer genes were mainly clustered in neurotransmitters as well as many chromatin assembly genes, both of which are important for neural plasticity and memory[36]. As transcription factors are increasingly recognized to bind to enhancers to achieve their regulatory roles, we performed a motif analysis of the cocaine-specific state 4 and state 5 putative enhancer regions and found substantial enrichment of a handful of candidate transcription factors (**Supplementary Table 5**). Some of these same transcription factors were deduced from our recent analysis of cocaine regulation of histone modifications and alternative splicing[33], which suggests that follow-up studies are necessary.

5hmC regulation of alternative splicing and transcription

Next, we explored the influence of altered 5hmC enrichment at coding regions of NAc in response to repeated cocaine. We identified enrichment of 5hmC changes at flanking exon boundaries (**Fig. 3a**), which implicates 5hmC in mRNA alternative splicing. Furthermore, we overlaid our 5hmC profiling with RNA-seq 24 h after repeated cocaine or saline treatment, the same time point used for our 5hmC experiments. We found enrichment of 5hmC-increased regions (35 regions) at splicing sites of upregulated splicing isoforms (1,488 sites), as well as enrichment of 5hmC-decreased regions (18 regions) at splicing sites of downregulated isoforms (1,536 sites; **Fig. 3b**,**c**). To the best of our knowledge, this is the first documentation of detailed alternative splicing sites coupled with 5hmC changes in brain (**Supplementary Table 6**). Together, our findings support the involvement of 5hmC at exonic regions in alternative RNA splicing in response to cocaine.

To complement our transcriptome analysis at 24 h of withdrawal from repeated cocaine treatment, we generated another RNA-seq data set, which was obtained 4 h after a subsequent challenge dose of cocaine (**Fig. 4a**). This data set better reflects altered inducibility of genes in contrast with the 24-h data set, which reflects longer lasting or steady-state changes in gene expression. Notably, cocaine-induced increases in 5hmC at gene bodies correlated significantly with increased gene expression after 24 h of withdrawal from cocaine exposure (odds ratio (observation/expectation) = 2.5, $P = 4 \times 10^{-3}$, 24 of 356 upregulated genes displayed increased 5hmC in gene body regions; **Fig. 4b**), and this overlap was even more significant with genes that were induced by a cocaine challenge (odds ratio = 3.5, $P = 10^{-6}$; odd ratio = ~1, $P > 0.05$ for all other cross comparisons; **Fig. 4b**). We observed a pre-existing increase in 5hmC enrichment in 31 of 261 genes upregulated at 4 h (**Table 1**). GO analysis indicated that these overlapping genes were concentrated in certain meaningful categories that have pivotal roles in drug addiction, such as long-term plasticity, synaptic transmission and glutamate neurotransmitter (**Table 1** and **Supplementary Fig. 7**). These indicate that repeated cocaine increased 5hmC correlates with both steady-state gene induction and gene inducibility in response to a subsequent cocaine challenge. Of note, however, the enhancer changes did not correlate genome-wide with gene expression (see Discussion).

To validate our findings, we carried out qPCR to measure gene expression and oxidative bisulfite sequencing (oxBS-seq) to quantify 5hmC alterations[37]. We first confirmed induction of several genes at the 24-h time point (*Adcy1*, *Itpr1*, *Hrk* and *Nsph4*; **Fig. 4c**) or 4 h

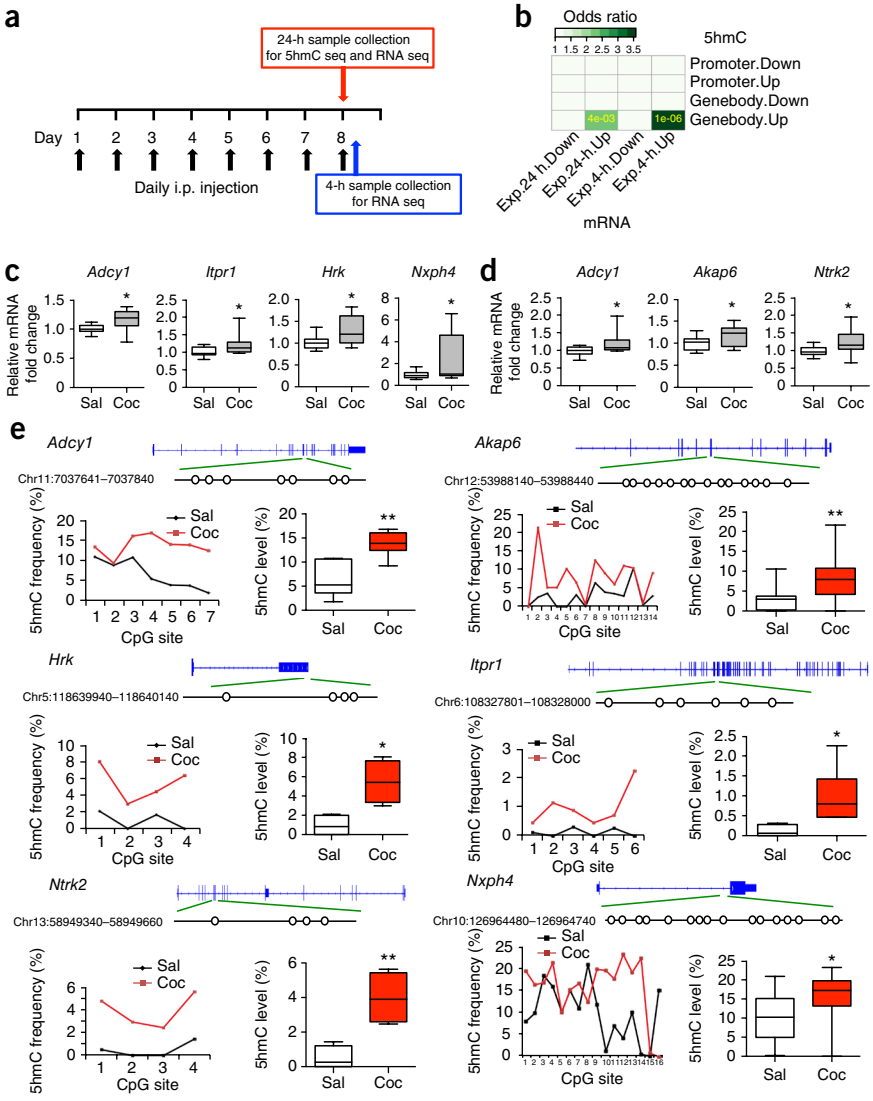

Figure 4 Gene body 5hmC alterations correlate with gene transcription changes after repeated cocaine. (**a**) Schematic drawing of the experimental timeline of 24-h 5hmC-seq and RNA-seq (24 h after a course of repeated cocaine treatment) and 4-h RNA-seq (4 h after a subsequent cocaine challenge). (**b**) Heat map showing significant overlaps between 5hmC alterations at gene bodies after repeated cocaine (24 h) and gene expression change (at 24 or 4 h). Color code represents odds ratio (observed/expected). Only significant *P* values are shown in grids. (**c**,**d**) qPCR validation (unpaired *t* test with Welch's correction except noted) of *Adcy1* ($P = 0.010$, $t(22) = 2.815$, $n = 12$ per group), *Itpr1* ($P = 0.043$, $t(22) = 2.151$, $n = 12$ per group), *Hrk* ($P = 0.030$, $t(18) = 2.345$, $n = 10$ per group) and *Nxph4* ($P = 0.041$, $t(21) = 2.177$, saline $n = 11$, cocaine $n = 12$ per group) 24 h after repeated cocaine (**c**) and *Adcy1* ($P = 0.035$, $t(12) = 2.384$, $n = 10$ per group), *Akap6* (unpaired *t* test, $P = 0.046$, $t(18) = 2.146$, $n = 10$ per group) and *Ntrk2* ($P = 0.019$, $t(12) = 2.719$, $n = 10$ per group) 4 h after a subsequent challenge (**d**). (**e**) oxBS-seq validated 5hmC differential changes (paired *t* test) in *Adcy1* ($P = 0.005$, $t(6) = 4.34$), *Akap6* ($P = 0.004$, $t(13) = 3.549$), *Hrk* ($P = 0.018$, $t(3) = 4.711$), *Itpr1* ($P = 0.034$, $t(5) = 2.888$), *Ntrk2* ($P = 0.005$, $t(3) = 7.692$) and *Nxph4* ($P = 0.034$, $t(15) = 2.324$). Schematic gene structures are illustrated on top. Each differential locus is enlarged with open circles denoting CpG sites. Genomic coordinate of the locus is shown. A summary of 5hmC frequency at each CpG site is demonstrated in a line chart and the mean 5hmC levels of individual alleles is displayed in a bar graph ($n = 2$–4 biological replicates per condition). Box plots present, in ascending order, minimum sample value, first quartile, median, third quartile, maximum sample value. $^*P < 0.05$, $^{**}P < 0.01$. Sal, saline; Coc, cocaine.

© 2015 Nature America, Inc. All rights reserved.

after a cocaine challenge (*Adcy1*, *Akap6* and *Ntrk2*; **Fig. 4d**) that were consistent with our RNA-seq findings. We then chose two genomic regions from our 5hmC-seq analysis for validation with oxBS-seq (**Supplementary Fig. 8**). This experiment confirmed induction of 5hmC at each of these genes in an independent cohort of animals (**Fig. 4e**).

To establish a causal role for the cocaine-induced TET1 downregulation in mediating the increased enrichment of 5hmC at gene targets, we investigated whether TET1 knockdown in NAc by itself in cocaine naive mice is able to mimic the 5hmC changes seen after cocaine. Indeed, for the six loci that we studied, five displayed robust induction of 5hmC following TET1 knockdown (**Fig. 5**).

Repeated cocaine-induced 5hmC regulation can be long lasting

5hmC is generally considered to be a transient intermediate state between 5mC and 5fC or 5caC that ultimately leads to unmethylated cytosine. However, 5hmC exists at much more abundant levels—particularly in brain—than 5fC or 5caC. Thus, we speculated that 5hmC might also serve as a stable epigenetic mark and tested whether some of the cocaine-induced changes in 5hmC persist long after cocaine exposure.

Of the genes we studied (**Fig. 4**), we demonstrated sustained induction of three of them (*Adcy1*, *Hrk* and *Ntrk2*; **Fig. 6**) 1 month after repeated cocaine treatment, and this induction was associated with increases in 5hmC (**Fig. 6**), effects not seen for the other genes (data not shown). These findings indicate that 5hmC can be a persisting epigenetic mark associated with prolonged transcriptional activation. Given that *Adcy1* and *Ntrk2* are important in drug addiction[38–40], the sustained induction of 5hmC might contribute to persistent cocaine actions in brain.

DISCUSSION

Our results reveal previously unappreciated roles for TET enzymes and 5hmC in the molecular and behavioral effects of cocaine. We observed selective downregulation of TET1 in NAc in response to chronic cocaine administration, which serves to promote behavioral responses to the drug. We found a link between such TET1 downregulation to alterations of 5hmC at putative enhancer regions and gene bodies, changes associated with altered gene expression and alternative splicing. Together, these findings reveal previously unknown epigenetic mechanisms that are involved in cocaine action.

Despite the fact that most studies of 5mC have focused on gene promoters, our finding that 5hmC alterations in gene bodies were associated with altered gene transcription is consistent with previous observations of 5hmC enrichment at gene bodies in stem cells[41]

Table 1 List of genes that show overlap of cocaine-induced changes by 5hmC-seq and RNA-seq

24-h genes	Gene ID	Representative GO terms
6330403A02Rik	ENSMUSG00000053963	Membrane
Adcy1	ENSMUSG00000020431	Adenyl nucleotide binding; learning or memory; behavior; long-term potentiation; Gap junction; memory; calcium signaling pathway; GnRH signaling pathway
Atp2b1	ENSMUSG00000019943	Calcium ion transport; phosphoprotein; ATP binding; brain development; aging
Crtac1	ENSMUSG00000042401	Calcium ion binding
Etl4	ENSMUSG00000036617	Glycoprotein; embryonic skeletal system development; glycosylation site; O-linked (GlcNAc); phosphoprotein
Fam49b	ENSMUSG00000022378	Breast cancer?
Foxo1	ENSMUSG00000044167	Phosphoprotein; nerve growth factor receptor signaling; transcription regulation
Hrk	ENSMUSG00000046607	Neuronal death; apoptosis
Hs3st4	ENSMUSG00000078591	Sulfotransferase activity
Itpr1	ENSMUSG00000030102	Long-term potentiation; synapse; metal ion transmembrane transporter activity; cellular ion homeostasis; calcium ion binding; phosphoprotein; GnRH signaling
Kremen1	ENSMUSG00000020393	Membrane; glycoprotein; nervous system development; Wnt signaling pathway
Lars2	ENSMUSG00000035202	ATP binding; gene expression; splicing; translation fidelity
Mapk10	ENSMUSG00000046709	GnRH signaling pathway; phosphoprotein; ATP binding; Reelin signaling in neurons; interleukin signaling; GDNF signaling
Mbp	ENSMUSG00000041607	Transmission of nerve impulse; cellular ion homeostasis; phosphoprotein
Ndrg3	ENSMUSG00000027634	Phosphoprotein; negative regulation of cell growth; positive regulation of Ras protein signaling
Nxph4	ENSMUSG00000040258	Neuropeptide signaling pathway
Pcsk5	ENSMUSG00000024713	Glycoprotein; cytoplasmic membrane-bounded vesicle; embryonic skeletal system development; nerve growth factor receptor signaling pathway
Rgnef	ENSMUSG00000021662	Plasma membrane; phosphoprotein; central nervous system neuron axonogenesis
Satb2	ENSMUSG00000038331	Embryonic skeletal system development; phosphoprotein
Scube1	ENSMUSG00000016763	Calcium ion binding; glycoprotein
Slc24a2	ENSMUSG00000037996	Memory; metal ion transmembrane transporter activity; cellular ion homeostasis; long term synaptic depression
Spock2	ENSMUSG00000058297	Calcium ion binding; phosphoprotein; glycoprotein; synapse assembly
Sv2b	ENSMUSG00000053025	Synapse; regulation of neurotransmitter levels; synaptic vesicle; transmission of nerve impulse; neurotransmitter secretion
Syn1	ENSMUSG00000037217	Transmission of nerve impulse; neurotransmitter secretion; ATP binding
4-h genes	**Gene ID**	**Representative GO terms**
Adcy1	ENSMUSG00000020431	Adenyl nucleotide binding; learning or memory; behavior; long-term potentiation; gap junction; memory; calcium signaling pathway; GnRH signaling pathway
Akap6	ENSMUSG00000061603	Spectrin/alpha-actinin; kinase; endomembrane system; nuclear envelope; positive regulation of cell growth
Arhgap32	ENSMUSG00000041444	Dendritic spine; cell membrane; phosphoprotein; dendrite; neuron projection; lipid binding; postsynaptic density; synapse
Atp1a3	ENSMUSG00000040907	Locomotory behavior; learning; cellular chemical homeostasis; ion homeostasis; ATP-binding; visual learning; visual behavior
Birc6	ENSMUSG00000024073	Apoptosis; protein ubiquitination
Cacna1e	ENSMUSG00000004110	Transport; calcium channel; di-, tri-valent inorganic cation transport; calcium signaling pathway; behavioral response to pain; fear response
Camk2a	ENSMUSG00000024617	Regulation of synaptic plasticity; autophosphorylation; phosphoprotein; GnRH signaling pathway; ATP binding; transcription factor CREB signaling; neurotransmitter secretion
Cyld	ENSMUSG00000036712	Phosphoprotein; central nervous system morphogenesis; apoptosis
D10Bwg1379e	ENSMUSG00000019852	Membrane; phosphoprotein
Daam2	ENSMUSG00000040260	Cytoskeletal protein binding; Wnt signaling pathway; coiled coil; actin binding; Wnt signaling pathway
Dst	ENSMUSG00000026131	Axonogenesis; calcium ion binding; cytoskeleton organization; intermediate filament-based process; regulation of microtubule polymerization or depolymerization; coiled coil
Gabrb1	ENSMUSG00000029212	Synapse; postsynaptic cell membrane; cell junction; ligand-gated channel activity
Gpr158	ENSMUSG00000045967	Phosphoprotein; G protein–coupled receptor signaling pathway
Gpr26	ENSMUSG00000040125	G protein–coupled receptor signaling pathway
Gria2	ENSMUSG00000033981	Dendrite; neuron projection; neurotransmitter receptor; glutamate receptor activity; synaptic transmission; transmission of nerve impulse; long-term depression; long-term potentiation
Grin2a	ENSMUSG00000059003	Synapse; memory; synaptic transmission; regulation of transmission of nerve impulse; calcium signaling pathway; glutamate receptor–related; learning or memory; behavior
Grin2b	ENSMUSG00000030209	Synapse; memory; synaptic transmission; regulation of transmission of nerve impulse; calcium signaling pathway; glutamate receptor–related; learning or memory; behavior
Hipk2	ENSMUSG00000061436	Locomotory behavior; phosphoprotein; ATP-binding; regulation of transcription
Ina	ENSMUSG00000034336	Coiled coil; cytoskeleton; phosphoprotein; nervous system development
Itpr1	ENSMUSG00000030102	Long-term potentiation; synapse; long-term potentiation; metal ion transmembrane transporter activity; cellular ion homeostasis; calcium ion binding; phosphoprotein; GnRH signaling
Kif1a	ENSMUSG00000014602	Axon guidance; anterograde synaptic vesicle transport; cytoskeleton; ATP-binding; coiled coil
Lars2	ENSMUSG00000035202	ATP binding; gene expression; splicing; translation fidelity
Mtap1b	ENSMUSG00000052727	Induction of synaptic plasticity; phosphoprotein; microtubule-based process; cytoskeleton
Ntrk2	ENSMUSG00000055254	Behavior; phosphoprotein; dendrite; ATP-binding; regulation of synaptic transmission; regulation of neurotransmitter secretion; regulation of transmission of nerve impulse; leaning and memory
Ppp1r16b	ENSMUSG00000037754	Coiled coil; signal transduction
Prkcb	ENSMUSG00000052889	Serine/threonine-specific protein kinase; long-term potentiation; calcium ion transport; synaptic long term depression; GnRH signaling pathway; NFkB signaling
Setd7	ENSMUSG00000037111	Histone lysine methylation
Sipa1l1	ENSMUSG00000042700	Coiled coil; regulation of synaptic plasticity; ephrin receptor signaling; Rap GTPase signaling
Syne1	ENSMUSG00000019769	Actin binding; brain development; dendrite morphogenesis
Tmtc1	ENSMUSG00000030306	Membrane
Unc13c	ENSMUSG00000062151	Coiled coil; synapse; cell junction; transmission of nerve impulse; synaptic transmission

The table lists those genes that show an increase both in 5hmc at gene bodies and in RNA expression at 4 or 24 h.

© 2015 Nature America, Inc. All rights reserved.

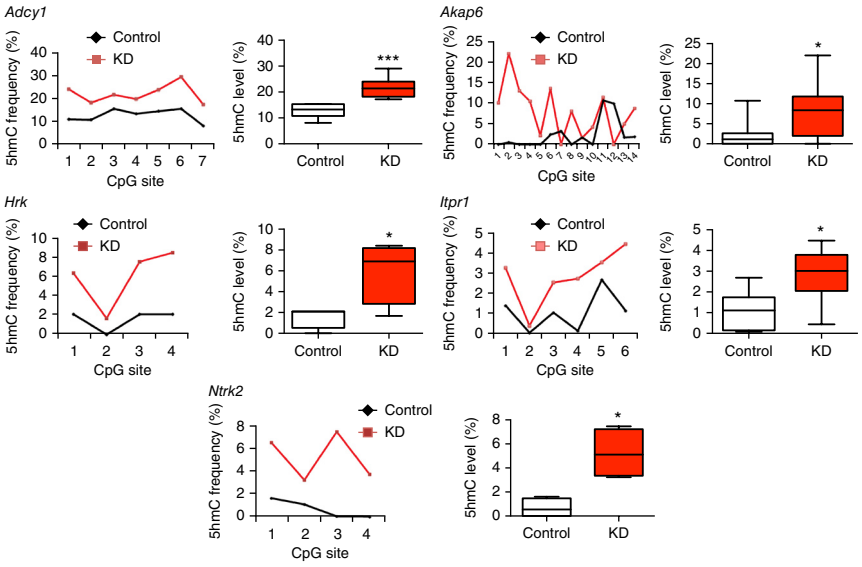

Figure 5 Tet1 knockdown in NAc of cocaine naive mice induces 5hmC at cocaine-regulated loci. oxBS-seq revealed increased 5hmC at *Adcy1* (paired *t* test, $P = 0.0002$, $t(6) = 8.017$), *Akap6* (paired *t* test, $P = 0.019$, $t(13) = 2.684$), *Hrk* (paired *t* test, $P = 0.022$, $t(3) = 4.356$), *Itpr1* (paired *t* test, $P = 0.011$, $t(5) = 3.945$) and *Ntrk2* (paired *t* test, $P = 0.027$, $t(3) = 4.082$) from NAc of naive animals that received viral *Tet1* shRNA knockdown as in **Figure 1e**. A summary of 5hmC frequency at each CpG site is demonstrated in a line chart and the mean 5hmC level of individual alleles is displayed in a bar graph ($n = 2$–4 biological replicates per condition). Box plots present, in ascending order, minimum sample value, first quartile, median, third quartile, maximum sample value. $^{*}P < 0.05$, $^{***}P < 0.001$. Control, control shRNA; KD, knockdown with *Tet1* shRNA.

and during neural development[19]. Although the exonic enrichment of 5hmC is suggestive of a role in pre-mRNA alternative splicing[42], our study is, to the best of our knowledge, the first to directly associate regulation at alternative splicing sites with altered splicing. We recently found that cocaine induces an order of magnitude more changes in alternative splicing than changes in total transcript levels[33]. Our results therefore implicate 5hmC in this prominent form of transcriptional regulation.

We also identified numerous cocaine-induced changes in 5hmC at putative enhancer sites. Using H3K4me1 and H3K27ac as tentative marks of active enhancers[28–32], we studied 5hmC dynamics at enhancer regions in response to cocaine. We found that 5hmC was enriched at enhancer regions in NAc and that such enrichment was regulated by cocaine. By assigning cocaine-regulated putative enhancers to the nearest genes to probe for potential enhancer targets, we found that the gene targets were enriched in interesting GO categories. However, we failed to detect a correlation between these regulated enhancers and our RNA-seq transcriptome profiling, emphasizing that the degree to which modifications at enhancer sites control dynamic regulation of gene expression in brain remains uncertain. Moreover, although defining enhancer-regulated targets by proximity alone has been a fruitful approach[29], we must emphasize its limitations. Many enhancers bypass hundreds of kilobases of interspersed sequence on the linear genome to interact with distant targets via chromosomal looping. Accordingly, studying the three-dimensional structure of chromatin can provide a precise prediction of enhancer target genes[43]. However, studies of higher order genome architecture are just beginning to be applied to the brain and major advances are needed before they can be applied to a microdissected brain region such as the NAc. Another means by which enhancers influence transcription is by directing the expression of a group of non-coding RNAs known as enhancer RNAs (eRNAs)[44] that are particularly sensitive to neural activity. However, many eRNAs are nonpolyadenylated[45], which is beyond the detection range of our poly-A selection–based Illumina RNA-seq protocol (Online Method). Moreover, given the cell type–specific feature of enhancers[30], enhancer-driven transcriptional regulation likely occurs in only a subtype of cells, which would be difficult to examine using our current heterogeneous tissue dissections and impossible to isolate in sufficient quantities for more selective analysis. Although functional implications of putative enhancers are often confirmed in *in vitro* reporter assays, this approach still requires further validation *in vivo*. Thus, continued technology development[43,46,47] will allow future investigations of enhancer regulation in NAc in addiction models. Nevertheless, despite these limitations, our

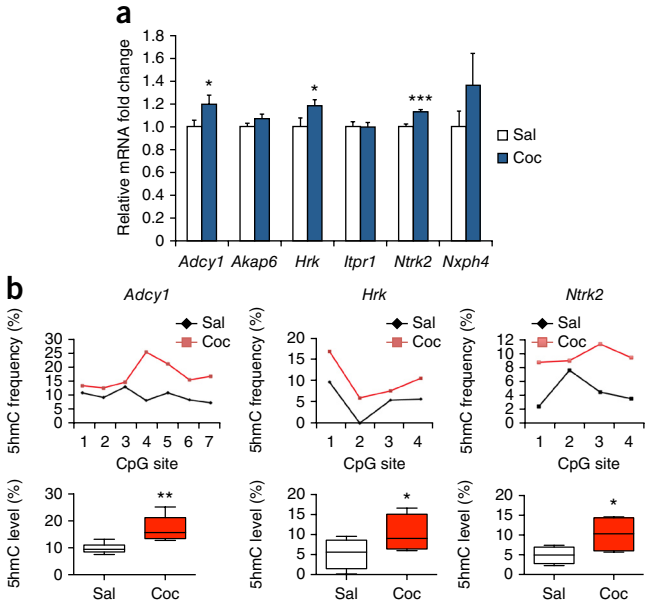

Figure 6 Long-lasting induction of 5hmC at particular loci after cocaine. (**a**) qPCR shows increased expression of *Adcy1* (unpaired *t* test, $P = 0.043$, $t(32) = 2.098$, $n = 17$ per group), *Hrk* (unpaired *t* test, $P = 0.033$, $t(30) = 2.242$, $n = 16$ per group) and *Ntrk2* (unpaired *t* test, $P = 7.04 \times 10^{-5}$, $t(30) = 4.607$, $n = 16$ per group) 1 month after repeated cocaine exposure, with no persisting change observed for *Akap6* (unpaired *t* test, $P = 0.140$, $t(30) = 1.517$, $n = 16$ per group), *Itpr1* (unpaired *t* test, $P = 0.944$, $t(30) = 0{,}071$, $n = 16$ per group) or *Nxph4* (unpaired *t* test, $P = 0.201$, $t(30) = 1.308$, $n = 16$ per group). Data are displayed as mean ± s.e.m. (**b**) oxBS-seq revealed concomitant increase of 5hmC at genes with persistent transcriptional induction (paired *t* test, *Adcy1*: $P = 0.012$, $t(6) = 3.567$; *Hrk*: $P = 0.018$, $t(3) = 4.688$; *Ntrk2*: $P = 0.027$, $t(3) = 4.050$). The results are displayed in the same manner as in **Figure 5**. $N = 2$–3 biological replicates per condition. Box plots present, in ascending order, minimum sample value, first quartile, median, third quartile, maximum sample value. $^{*}P < 0.05$, $^{**}P < 0.01$. Sal, saline; Coc, cocaine.

© 2015 Nature America, Inc. All rights reserved.

putative identification of enhancer regions suggests the involvement of 5hmC in enhancer state dynamics. Our analysis also identified several transcription factor motifs at cocaine-regulated putative enhancer regions, which is consistent with the emerging role of transcription factors in non-coding enhancer regions[29].

Our finding that cocaine-induced 5hmC enrichment in NAc positively correlates with increased gene expression is consistent with previous reports[48]. 5hmC may promote such transcription activation through dissociation of 5mC binding protein repressors or recruitment of previously unknown effector proteins[15]. Previous research proposed that a decrease in TET1 leads to less 5mC conversion to 5hmC, and thus to promoter hypermethylation and transcription repression[20–22]. Our demonstration that TET1 downregulation leads to a substantial overlap between increased 5hmC and gene induction after cocaine appears to counter this view. There are several possible explanations. First, the previous studies focused on gene promoters, whereas the alterations in 5hmC that we identified occurred at gene bodies, and it is known that DNA modifications exert different effects at these distinct regions[49]. Second, none of the earlier studies compared 5hmC landscapes with genome-wide transcriptome profiling, as we did, and instead used candidate gene approaches. Third, although 5mC oxidation patterns were not examined in some studies[20,21], TET1 knockout mice did display elevated 5hmC at selected genes[22], suggesting that TET1 deficiency can cause increased 5hmC, at least at certain loci. Fourth, if TET1 downregulation causes universal DNA hypermethylation, it is hard to explain why there is a global decrease in 5mC when TET1 is repressed in mouse hippocampus in the context of seizures[21]. In fact, TET1 has been implicated in both transactivation and repression[41,50]. In TET1-depleted embryonic stem cells, more TET1 targets were upregulated than downregulated. It is hypothesized that this TET1 effect might reflect its normal coupling with repressor complexes[41,50]. Moreover, how 5hmC conversion into 5fC or 5caC is affected by TET1 deficiency is unknown. An interruption of this process after TET1 loss might contribute to 5hmC accumulation. Notably, we found that TET1 knockdown in NAc caused similar induction of 5hmC (**Fig. 5**), as observed after cocaine exposure, which supports a causal role of TET1 downregulation in 5hmC induction at the genomic loci that we investigated. However, the lack of a global DNA methylation change (**Fig. 2a**), the observation that TET1 knockdown did not mimic all cocaine-triggered 5hmC increases in NAc and possible non-enzymatic activities of TET1 (ref. 21) suggest several directions for future research. In addition, it would be interesting to explore a possible role for TET2 and TET3 in cocaine action, even though they are not regulated in NAc by cocaine, given their relative abundance in this brain region (**Supplementary Fig. 3**). Nevertheless, our data indicate that TET1 has at least some unique targets in NAc that are not compensated by TET2 or TET3 and that demonstrate 5hmC induction in response to cocaine.

It has been challenging to correlate DNA epigenetic modifications and gene expression regulation genome-wide in brain under several experimental procedures. This is possibly because available methodologies are not sufficiently sensitive to capture minor changes of both simultaneously, most studies have focused on a subset of genomic regions (for example, promoters) and might have missed important regulation elsewhere, sodium bisulfite sequencing technology used to date for DNA methylation analysis do not differentiate 5hmC and 5mC, which can exert opposite effects on transcription, and a change in DNA methylation can reflect not only an alteration in steady-state gene expression, but also a gene's poising for future transcription. Our study addresses several of these questions, as we found, genome-wide in a discrete region of adult brain, that dynamic regulation of 5hmC correlated with several features of gene expression, including alternative splicing of primary transcripts, alterations in steady-state levels of expression and future inducibility, and that some of the 5hmC changes were long lasting.

In summary, our results establish 5hmC as a previously unknown epigenetic mark in cocaine action. By mapping 5hmC in NAc in response to repeated cocaine exposure, we identified numerous target genes that now warrant investigation for their role in cocaine action. That some of the genes identified have previously established roles in cocaine addiction[46–48] supports the importance of examining the targets. A prominent feature of drug addiction is that, once it is formed, those addicted can have life-long behavioral abnormalities, which implicates stable brain changes in this process[24]. Our findings present 5hmC as one plausible underpinning mechanism. The next important steps are to carry out more sophisticated behavioral experiments of drug addiction, such as self-administration, to probe a role of 5hmC at particular genomic loci in NAc in the development and maintenance of addiction. Such work has the potential of generating new means of interrupting this process for therapeutic goals.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. All new ChIP-seq and RNA-seq data have been deposited into the Gene Expression Omnibus with accession number GSE63749. Datasets of 24-h RNA-seq and H3K4me1 and H3K4me3 ChIP-seq were previously deposited with accession number GSE42811.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank A. Chess for critical comments and O. Jabado and M. Mahajan from the Mount Sinai Genomics Core for technical support. This work was supported by grants from the National Institute on Drug Abuse (E.J.N.), the National Institutes of Health (P.J., P.C., G.F., K.F.F. and H.S.) and the Simons Foundation (H.S.).

AUTHOR CONTRIBUTIONS

The studies were conceived and designed by E.J.N., P.J. and J.F. J.F. performed RNA-seq and ChIP-seq. K.E.S., Y.L. and J.F. performed 5hmC capture and sequencing. L.S., N.S. and J.F. performed bioinformatic analyses. J.F. and J.H. performed oxBS-seq. JF., V.V., B.M.L., V.S. and I.M. performed qPCR analyses. V.V., D.E., T.C. and J.F. performed immunohistochemistry. J.F., V.V., D.F., J.K. and E.R. performed stereotaxic surgeries and behavioral assays. T.L., K.F.F. and G.F. contributed LC-ESI-MS/MS data. C.Z., G.M. and H.S. provided AAV-Tet1 shRNA and AAV-TET1 viruses. M.E.C. performed western blotting. P.C. contributed reagents. G.T. contributed human samples. D.F., B.L., B.M.L., V.S. and P.K. helped to prepare the samples and collect the data. The paper was written by J.F. and E.J.N. and was edited by the other authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33** (suppl.), 245–254 (2003).
2. Weaver, I.C. *et al.* Epigenetic programming by maternal behavior. *Nat. Neurosci.* **7**, 847–854 (2004).
3. Miller, C.A. & Sweatt, J.D. Covalent modification of DNA regulates memory formation. *Neuron* **53**, 857–869 (2007).
4. Ma, D.K. *et al.* Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science* **323**, 1074–1077 (2009).
5. Feng, J. *et al.* Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.* **13**, 423–430 (2010).
6. Ooi, S.K. & Bestor, T.H. The colorful history of active DNA demethylation. *Cell* **133**, 1145–1148 (2008).

© 2015 Nature America, Inc. All rights reserved.

7. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
8. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
9. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
10. Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).
11. He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
12. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
13. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* **25**, 2436–2452 (2011).
14. Branco, M.R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2012).
15. Pastor, W.A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14**, 341–356 (2013).
16. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
17. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
18. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181 (2010).
19. Szulwach, K.E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14**, 1607–1616 (2011).
20. Rudenko, A. *et al.* Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**, 1109–1122 (2013).
21. Kaas, G.A. *et al.* TET1 controls CNS 5-methylcytosine hydroxylation, active DNA Demethylation, gene transcription and memory formation. *Neuron* **79**, 1086–1093 (2013).
22. Zhang, R.R. *et al.* Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell* **13**, 237–245 (2013).
23. Li, X. *et al.* Neocortical Tet3-mediated accumulation of 5-hydroxymethylcytosine promotes rapid behavioral adaptation. *Proc. Natl. Acad. Sci. USA* **111**, 7120–7125 (2014).
24. Nestler, E.J. Molecular basis of long-term plasticity underlying addiction. *Nat. Rev. Neurosci.* **2**, 119–128 (2001).
25. Feng, J. & Nestler, E.J. Epigenetic mechanisms of drug addiction. *Curr. Opin. Neurobiol.* **23**, 521–528 (2013).
26. LaPlant, Q. *et al.* Dnmt3a regulates emotional behavior and spine plasticity in the nucleus accumbens. *Nat. Neurosci.* **13**, 1137–1143 (2010).
27. Le, T., Kim, K.P., Fan, G. & Faull, K.F. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. *Anal. Biochem.* **412**, 203–209 (2011).
28. Rajagopal, N. *et al.* RFECS: a random-forest based algorithm for enhancer identification from chromatin state. *PLOS Comput. Biol.* **9**, e1002968 (2013).
29. Malik, A.N. *et al.* Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat. Neurosci.* **17**, 1330–1339 (2014).
30. Heintzman, N.D. *et al.* Histone modifications at human enhancers reflect global cell type–specific gene expression. *Nature* **459**, 108–112 (2009).
31. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936 (2010).
32. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
33. Feng, J. *et al.* Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. *Genome Biol.* **15**, R65 (2014).
34. Sasai, N., Nakao, M. & Defossez, P.A. Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Res.* **38**, 5015–5022 (2010).
35. Huh, G.S. *et al.* Functional requirement for class I MHC in CNS development and plasticity. *Science* **290**, 2155–2159 (2000).
36. Vogel-Ciernia, A. *et al.* The neuron-specific chromatin regulatory subunit BAF53b is necessary for synaptic plasticity and memory. *Nat. Neurosci.* **16**, 552–561 (2013).
37. Booth, M.J. *et al.* Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**, 934–937 (2012).
38. Zachariou, V. *et al.* Distinct roles of adenylyl cyclases 1 and 8 in opiate dependence: behavioral, electrophysiological and molecular studies. *Biol. Psychiatry* **63**, 1013–1021 (2008).
39. Graham, D.L. *et al.* Tropomyosin-related kinase B in the mesolimbic dopamine system: region-specific effects on cocaine reward. *Biol. Psychiatry* **65**, 696–701 (2009).
40. Lobo, M.K. *et al.* Cell type–specific loss of BDNF signaling mimics optogenetic control of cocaine reward. *Science* **330**, 385–390 (2010).
41. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
42. Khare, T. *et al.* 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat. Struct. Mol. Biol.* **19**, 1037–1043 (2012).
43. Mitchell, A.C. *et al.* The genome in three dimensions: a new frontier in human brain research. *Biol. Psychiatry* **75**, 961–969 (2014).
44. Kim, T.K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
45. Natoli, G. & Andrau, J.C. Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* **46**, 1–19 (2012).
46. Pennacchio, L.A. *et al.* *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
47. Cruz, F.C. *et al.* New technologies for examining the role of neuronal ensembles in drug addiction and fear. *Nat. Rev. Neurosci.* **14**, 743–754 (2013).
48. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417–1430 (2012).
49. Hellman, A. & Chess, A. Gene body–specific methylation on the active X chromosome. *Science* **315**, 1141–1143 (2007).
50. Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389–393 (2011).

© 2015 Nature America, Inc. All rights reserved.

ONLINE METHODS

Animals and cocaine administration. Adult male C57BL/6J mice (Jackson Lab) 8–10 weeks old were used in this study. They were housed four to five per cage on a 12-h light-dark cycle at constant temperature (23 °C) with free access to food and water *ad libitum*. All animals were experimentally naive. They were habituated for at least 1 week before experimentation and assigned randomly to experimental groups. All mice were from at least three litters in each group. All behavioral experiments were performed during the light cycle. All animal protocols were approved by the Institutional Animal Care and Use Committee of Mount Sinai. All experiments were carried out by experimenters blind to animal treatment conditions whenever possible (this includes all behavioral assays, RNA/DNA isolation, IHC and western blotting).

Animals received daily intraperitoneal injections for 7 consecutive days of cocaine (Sigma) at 20 mg per kg body weight (repeated cocaine). Mice were used 24 h after the final injection. For the 4-h experiments, mice were killed 4 h after a challenge dose of cocaine (20 mg per kg, intraperitoneal) administered 24 h after repeated cocaine. Control mice for all groups received saline injections. Bilateral 14-gauge NAc punches were taken from each animal. Unless otherwise noted, NAc punches from five mice were pooled for one replicate in ChIP assay, whereas NAc punches from one mouse were used as one replicate for all other RNA/DNA/protein assays.

Human samples. NAc samples from humans addicted to cocaine were acquired from the brain bank at McGill University. Brain tissues were from addicted subjects and healthy control male adult subjects of French Canadian ancestry. The cohort was composed of 37 male subjects, ranging in age between 15 and 66 years. All subjects died suddenly without a prolonged agonal state or protracted medical illness. In each case, the cause of death was ascertained by the Quebec Coroner Office, and a toxicological screen was conducted with tissue samples to obtain information on medication and illicit substance use at the time of death. The subject group consisted of individuals who met the Structured Clinical Interview for DSM-IV (Diagnostic and Statistical Manual of Mental Disorders-IV) Axis I Disorders: Clinician Version (SCID-I) criteria for cocaine dependence. The control group comprised subjects with no history of cocaine dependence and no major psychiatric diagnoses. All subjects died suddenly from causes that had no direct influence on brain tissue. Groups were matched for mean subject age, refrigeration delay, and pH. For all subjects, psychological autopsies were performed as described previously (see ref. 51 for further details on patient and control demographics).

RNA isolation and qPCR. Brain samples were homogenized in Trizol (Invitrogen) and processed according to the manufacturer's instructions. RNA was purified with RNeasy Micro columns (Qiagen) and spectroscopy confirmed that the RNA 260/280 rations were >1.8. RNA was then reverse transcribed using a Bio-Rad iScript Kit (Biorad). cDNA was quantified by qPCR using SYBR Green (Quanta). Each reaction was run in duplicate or triplicate and analyzed following the ΔΔCt method as previously described[5] using glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and B2M as normalization control. All experiments were repeated at least twice. All primer sequences are listed in **Supplementary Table 7**.

Immunohistochemistry. Coronal sections 25 μm thick were prepared on a microtome and were processed with a regular free floating immunohistochemistry protocol. After 3 washes in phosphate-buffered saline and a blocking with 4% (wt/vol) BSA, 0.4% (vol/vol) Triton X-100 in phosphate-buffered saline, brain sections were incubated with antibodies to TET1 (Millipore cat# 09-872, 1:500), H1 (Abcam cat# ab62884, 1:2,000), DARP32 (Cell signaling cat# 2306, 1:1,000), or GFP (Molecular Probes cat# A11120, 1:1,000) at 4 °C overnight. They were then incubated with secondary antibodies: Alexa 488– or Alexa 647–conjugated secondary antibody (Molecular Probes) at room temperature (23–25 °C) for 2 h. Quantification of TET1 immunoreactivity was also determined with a Licor system by using Odyssey software as previously described[52]. Results are presented as integrated intensity values per mm^2 and are presented as means ± s.e.m. Values for histone linker protein H1 were used as normalization controls. Ratios of TET1 over total H1 were analyzed, and Student's *t* test was used to compare means for each brain region. Differences were considered significant at value of $P < 0.05$. These IHC experiments were successfully repeated at least once.

Western blot analysis. The nucleus accumbens was dissected from mice and subjected to a subcellular fraction protocol that produces multiple cell fractions, including a nuclear fraction, using established methods[53]. Specifically, mouse nucleus accumbens was solubilized in HEPES-buffered sucrose (0.32 M sucrose, 4 mM HEPES, with protease and phosphatase inhibitors, pH = 7.4) using a teflon homogenizer at 400 rpm and approximately 40 up and down strokes. The solution then was centrifuged at 1,000 *g* at 4 °C for 10 min. The resulting pellet (p1 fraction) was isolated and resuspended in HEPES-buffered sucrose and again centrifuged at 1,000 *g* at 4 °C; this step was repeated three times. Next, the p1 pellet was resuspended in radio-immunoprecipitation buffer (RIPA, pH = 7.4). The resulting purified p1 fraction is enriched for cell nuclei. Protein concentration was determined using a DC protein assay (Bio-Rad). Equal amounts of protein were then mixed with 5× sample loading buffer and boiled at 95°C for 5 min. Samples were separated using SDS-PAGE, and protein levels quantified using densitometry (Image J). All western blot analyses were successfully repeated at least once. Antibodies to TET1 (Millipore cat# 09-872), TET2 (Santa Cruz, cat# sc-136926) and TET3 (Millipore cat# ABE383) were used in this experiment.

Virus and stereotaxic surgery. AAV-*TET1* overexpression and AAV-GFP control vectors were the same as reported previously[16]. *Tet1* shRNA lentiviral particles (cat# sc-154204-V) and control shRNA lentiviral particles (cat# sc-108080) were purchased from Santa Cruz Biotechnology. The second *Tet1* knockdown AAV-Tet1 shRNA construct was designed following published literature[9]. Stereotaxic intra-NAc injection was performed according to published protocols[52]. Viral injection sites were verified by confirming the GFP signal in NAc slices under the microscope. Animals with incorrect injection placement were excluded from analyses. Viral manipulation of *Tet1* was confirmed using qPCR and western blot.

CPP. A standard, unbiased CPP procedure was used as previously described[33]. In brief, 3–4 weeks after viral injection, animals were pretested for 20 min in a photo-beam monitored box with free access to environmentally distinct chambers. The mice were then arranged into control and experimental groups with balanced pretest scores. Mice underwent four 30-min training sessions (saline morning and cocaine in the afternoon) over 2 d. On the test day, mice had 20 min of unrestricted access to all chambers and a CPP score was assigned by subtracting the time spent in the cocaine-paired chamber from the time spent in the saline-paired chamber. Cocaine was injected intraperitoneal at 7.5 mg per kg.

5mC and 5hmC quantification. Genomic DNA was isolated from NAc punches by using Qiagen's DNA purification kit. The liquid chromatography–electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) quantification of 5mC and 5hmC was performed with established protocols[5,27].

5hmC capture and sequencing. 5hmC enrichment and sequencing were performed with a previously described selective chemical labeling method[19,54]. Genomic DNA was first sonicated into 100–500 bp by using a Covaris instrument. 5hmC labeling reactions were performed in a 100 μl solution containing 50 mM HEPES buffer (pH 7.9), 25 mM $MgCl_2$, sonicated DNA, 250 μM UDP-6-N3-Glu, and 2.25 μM wild-type β-glucosyltransferase. DNA substrates were then purified via DNA purification kit (Qiagen). Click chemistry was performed with the addition of 150 μM dibenzocyclooctyne modified biotin into the DNA solution, and incubated at 37 °C for 2 h. Samples were then purified with Pierce Monomeric Avidin Kit (Thermo). After elution, biotin-5-N3-gmC–containing DNA was concentrated and purified using a Qiagen DNA purification kit.

Libraries were generated following the Illumina protocol for "Preparing Samples for ChIP Sequencing of DNA" (cat# 111257047 Rev. A) with 25 ng 5hmC–captured DNA. DNA fragments of ~150–300 bp were size selected through agarose gel-purification after adaptor ligation. PCR-amplified DNA libraries were quantified on Agilent Bioanalyzer and diluted to 6–8 pM for cluster generation and sequencing. 38-cycle single-end sequencing was carried out by use of a Version 4 Cluster Generation (cat# 15002739) and Sequencing Kits and (cat# 15005236) and Version 7.0 recipes. Image processing and sequence extraction were accomplished using the standard Illumina pipeline. Three biological replicates were performed for each condition.

mRNA sequencing. 4 μg of total RNA was used for mRNA library construction following instructions of Illumina mRNA sample prep kit (cat# RS-100-0801).

© 2015 Nature America, Inc. All rights reserved.

In brief, the poly-A–containing mRNA was purified using poly-T oligo-attached magnetic beads. The mRNA was then fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers. This was followed by second strand cDNA synthesis using DNA Polymerase I and RNaseH. These cDNA fragments went through an end repair process, the addition of a single 'A' base, and then ligation of the adapters. These products were gel purified and enriched with PCR to create the final cDNA libraries. The library constructs were run on the bioanalyzer to verify the size and concentration before sequencing on the Illumina HiSeq2000 machine at Mount Sinai's Genomic Core facility. In total, three biological replicates were carried out for each condition.

ChIP and ChIP-seq. ChIP was performed as previously described[55]. 3–5 µg ChIP grade antibodies to H3K4me1 (Abcam cat# ab8895), H3K4me3 (Abcam cat# ab8580), H3K27ac (Abcam cat# ab4729) or P300 (Santa Cruz cat# sc-585 X) were used per ChIP reacion. Around 10 ng of pull-down DNA or input DNA was used for preparation of sequencing libraries following the instructions of Illumina's ChIP-seq sample prep kit (cat# IP-102-1001). In brief, DNA fragment overhangs were converted into phosphorylated blunt ends, using T4 DNA polymerase, Klenow polymerase, and T4 polynucleotide kinase. An 'A' base was then added to the DNA fragments to enable ligation to the adapters, which have a single 'T' overhang. The constructed library was run on a 2% agarose gel and size selected between 175–300 bp. Lastly, gel-extracted DNA was further enriched by PCR and run on a bioanalyzer to validate size distribution and concentration. All sequencing libraries were run on Illumina Hi-seq 2000 machines. Three biological replicates were done for each condition.

oxBS sample preparation and analysis. DNA samples were oxBS converted with the TrueMethyl kit (Cambridge Epigenetix) or as previously reported with slight modifications[37,56]. In brief, genomic DNA was isolated from NAc tissues pooled from 2–3 animals as one replicate. Two µg of DNA was precipitated and purified four times with Mini Quick Spin Oligo Columns (Roche cat# 11814397001) and eluted in 30 µl. Two thirds of the DNA (20 µl) was aliquoted and 4 µl of 0.3 M NaOH was added. After incubation at 37 °C for 30 min, 1 µl $KRuO_4$ (15 mM in 0.05 M NaOH) was added and left on ice for 1 h with occasional vortex. The $KRuO_4$-treated DNA was purified again with mini quick spin oligo column. The remaining one third purified DNA (10 µl) and the $KRuO_4$-treated DNA were processed through sodium bisulfite conversion by using Zymo EZ DNA methylation kit (cat# D5001) and identified as BS DNA and oxBS DNA, respectively. BS DNA and oxBS DNA were then used for PCR with primers designed flanking target region of interest. PCR amplicons of various primer pairs were then purified and pooled together as for library preparation. Library preparation followed NEB Next DNA Library Prep kit (cat# 6040) with a barcode for each sample. The libraries were examined on a bioanalyzer to determine concentration and size. All libraries were then pooled with equal concentration to get the final library mixture and processed for Illumina MiSeq 250 bp pair end sequencing. The raw sequencing reads were processed by Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), and the low sequencing quality nucleotides (quality score <30) at the 3′ were removed, and only the reads longer than 80 bp were retained. Alignment and methylation calling of the filtered reads were then performed with Bismark (http://www.bioinformatics.babraham.ac.uk/projects/bismark/)[57]. Samples with low read depth (<50 reads/sample) were removed. The unconverted CpG ratio was calculated for each CpG site as unconverted read counts/total read counts. 5hmC frequency at each CpG site was derived as unconverted CpG ratio in BS-seq subtracts the counterpart in oxBS-seq of the same DNA sample. 5hmC ratio was then determined as the average of all CpG sites in the same allele. The Student's *t* test (two-tailed) was used for statistical analysis. Statistical significance was determined at $P < 0.05$.

Statistical analysis. Sample sizes were similar to those reported in previous works and based on expected effect sizes. Data were processed and analyzed with Prism 5.0 (GraphPad). Normality and equal variance assumptions were tested through Prism before parametric analyses. For assessments of differences between two experimental groups, student's unpaired *t* test (two tailed) was used, except a paired *t* test was applied for ox-BS analyses. For comparisons with unequal variances, Student's *t* test with Welch's correction was used as noted. Outliners were excluded from analysis when identified by Grubbs' test, which only happened occasionally. Differences were considered significant at value of $P < 0.05$.

5hmC-seq and ChIP-seq differential analysis. Short reads were first aligned to the mouse reference genome (mm9) using the Bowtie[58] short read aligner. The redundantly aligned reads were removed to eliminate potential PCR bias in amplification. diffReps[59] was then used to detect 5hmC differential regions with a window size of 200 bp and a moving size of 20 bp to get a high-resolution profiling of the 5hmC changes regulated by cocaine. diffReps uses a sliding window approach to scan the genome and performs statistical inferences using the negative binomial test. It calculates a *P* value for each window and then uses the Benjamin-Hochberg method to control for false discovery rate. In this study, a *P* value of 10^{-3} was used for identifying candidate windows. An FDR of 15% was used to determine the final list of differential regions.

Enhancer analysis. H3K4me1, H3K4me3, H3K27ac, and 5hmC ChIP-seq data were used for detection of enhancers. All ChIP-seq data sets were aligned to the mouse genome (mm9) by Bowtie[58]. Based on chromatin combinations among the genome, ChromHMM (http://compbio.mit.edu/ChromHMM/) implemented a multivariate hidden Markov model (HMM) to investigate chromatin state[60]. Here we used the above mentioned histone marks to investigate the chromatin dynamic states with ChromHMM, using 200-nt length window size[32]. H3K4me1 and H3K27ac enriched sites were treated as putative active enhancers, while H3K4me3 and H3K27ac enriched sites were treated as putative promoters. Based on chromatin states, enhancers were further classified based 5hmC enrichment (two enhancers with a distance of <500 nt were merged). Average profile plots and heatmaps of the histone marks across the genome were plotted by ngs.plot[61]. Enhancers were annotated by HOMER[62], and the nearest genes to the enhancers were extracted. Gene lists for cocaine- and saline-treated conditions were compared, and the condition-specific enhancers were kept for GO analysis. For motif analysis, sequences of putative enhancers and the flanking regions (±1 kb) were extracted and the motif enrichment examination were done by MEME package[63].

GO analysis. All GO analyses were performed by DAVID (http://david.abcc.ncifcrf.gov/)[64] with default settings except using high stringency for Functional Annotation Clustering analysis.

RNA-seq analysis. Short read alignment, transcript expression estimation, and differential analysis were carried out using the Tophat and Cufflinks pipelines[65]. The short reads were aligned against a reference gene database (Ensembl: *Mus musculus*, NCBIM37.62) without attempting to assemble novel transcripts. Both bias detection (–frag-bias-correct) and multi-read correction (–multi-read-correct) options were used to obtain a more accurate estimate of transcript abundance. A gene is considered differentially expressed with FDR < 15%. The RNA splicing is analyzed by Cuffdiff with cutoff at FDR < 10%. A custom Perl script is used to extract the differentially spliced exons from the Cuffdiff results. In total, there are 1,488 upregulated exons and 1,536 downregulated exons.

Correlation analysis between 5hmC differential regions and gene expression. The 5hmC differential regions were mapped to genes according to their proximity to a gene body. According to its distance to the transcriptional start site (TSS), the differential regions were further classified into two categories: promoter (±3 kb) and genebody (>+3 kb). Gene groups were formed by the locations of the differential regions that they contained. Pair-wise comparisons between differential region and gene expression groups were performed using Fisher's exact test using the whole mouse genome as universe. *P* values were further adjusted by the BH method[66]. Any grid containing $P > 0.01$ is considered non-significant and its odds ratio is not shown.

Enrichment of 5hmC differential regions at alternatively splicing sites. The numbers of 5hmC differential regions at alternatively splicing sites were determined by overlapping genomic locations allowing a 500 bp gap, for both directions separately. Pairwise comparisons were performed for the three groups of splicing transcripts with increased, decreased, and non-significant changes using Fisher's exact test.

© 2015 Nature America, Inc. All rights reserved.

doi:10.1038/nn.3976

Coverage plots. Coverage plots at TSS, TES, and genebody were generated using ngs.plot[61]. Briefly, we determined the read count for each base-pair and then normalized them by the total library size to get the coverage, or the so-called 'read count per million mapped reads' (RPM). These coverage vectors were then averaged across all the functional regions, such as all TSSs, on the genome to obtain an average profile. For gene bodies, where regions are of variable length, a spline fit was first performed and then sampled at equal intervals so that all coverage vectors were normalized to the same length.

A **Supplementary Methods Checklist** is available.

51. Robison, A.J. *et al.* Behavioral and structural responses to chronic cocaine require a feedforward loop involving DeltaFosB and calcium/calmodulin-dependent protein kinase II in the nucleus accumbens shell. *J. Neurosci.* **33**, 4295–4307 (2013).
52. Vialou, V. *et al.* DeltaFosB in brain reward circuits mediates resilience to stress and antidepressant responses. *Nat. Neurosci.* **13**, 745–752 (2010).
53. Li, N. *et al.* mTOR-dependent synapse formation underlies the rapid antidepressant effects of NMDA antagonists. *Science* **329**, 959–964 (2010).
54. Song, C.X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29**, 68–72 (2011).
55. Vialou, V. *et al.* Serum response factor and cAMP response element binding protein are both required for cocaine induction of DeltaFosB. *J. Neurosci.* **32**, 7577–7584 (2012).
56. Booth, M.J. *et al.* Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat. Protoc.* **8**, 1841–1851 (2013).
57. Krueger, F. & Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
58. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
59. Shen, L. *et al.* diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS ONE* **8**, e65598 (2013).
60. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
61. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).
62. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
63. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
64. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
65. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
66. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

© 2015 Nature America, Inc. All rights reserved.