5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary

Tarang Khare1,15, Shraddha Pali1,15, Karolis Koncevicius1,2, Mrinal Pal1, Edita Kriukiene3, Zita Liutkeviute3, Manuel Irimia4, Peixin Jia1, Carolyn Ptak1, Menghang Xia5, Raymond Tice6, Mamoru Tochigi7, Solange Moréra8, Anaïes Nazarians9,10, Denise Belsham9,10, Albert H C Wong11, Benjamin J Blencowe4, Sun Chong Wang1,12, Philipp Kapranov13, Rafal Kustra14, Viviane Labrie1, Saulius Klimasauskas3 & Arturas Petronis1

The recent rediscovery of 5-hmC in mammals1–3 demonstrates that covalent DNA modifications are more dynamic than previously believed. The oxidation of 5-mC generates 5-hmC in a reaction mediated by the ten-eleven translocation (TET) family of enzymes (TET1, TET2 or TET3)3. Traditional means of assaying DNA modifications either have been unable to differentiate between 5-mC and 5-hmC (for example, bisulfite mapping4) or have been specific for 5-mC (for example, antibodies against 5-mC). 5-hmC is particularly enriched in the brain relative to other tissues, as observed in mice and humans1,5. At gene promoters, 5-mC suppresses transcription by recruiting transcriptional repressors6. In the mouse cortex and cerebellum, 5-hmC is enriched within genes and appears to increase with increasing transcription levels7,8. In addition to being an intermediate state of complete DNA demethylation, 5-hmC also effects functional demethylation by suppressing binding of transcriptional repressors9–12. Synchronous neuronal activity promotes active DNA demethylation of plasticity-related genes in the mouse brain through TET-mediated formation of 5-hmC (ref. 9); however, it is not known whether such demethylation completely accounts for the enrichment of 5-hmC in the brain.

We mapped both 5-mC and 5-hmC in a variety of neuronal and non-neuronal tissues from mice and humans to investigate their respective roles. We labeled 5-hmC using the phage enzyme β-glucosyltransferase (BGT), followed by differential digestion of DNA with restriction enzymes either sensitive or insensitive to these DNA modifications (Fig. 1a). The resulting DNA fragments were amplified and analyzed by genome-wide tiling microarray analysis. We detected 5-hmC enrichment in genes with synapse-related spliced exons. Our study suggests a new role for 5-hmC in RNA splicing and synaptic function in the brain.

RESULTS
Validation of 5-hmC assay
BGT transfers a glucose molecule specifically to the hydroxymethyl group of 5-hmC, thus rendering it resistant to digestion by the methylation-insensitive MspI enzyme at the CHmCGG target site13,14 (Supplementary Fig. 1a). Thus, 5-hmC abundance is estimated by differential resistance to MspI digestion with and without glucosylation of genomic DNA (gDNA). HpaII targets the same site, CCGG, yet it cannot cut CHmCGG or CHmCCG, so comparing HpaII digestion with MspI digestion provides a measure of both 5-mC and 5-hmC abundance. Subtraction of the 5-hmC estimate

1The Krembil Family Epigenetics Laboratory, The Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada. 2Institute of Mathematics and Informatics, Vilnius University, Vilnius, Lithuania. 3Department of Biological DNA Modification, Institute of Biotechnology, Vilnius University, Vilnius, Lithuania. 4The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada. 5National Center for Advancing Translational Sciences, US National Institutes of Health, Bethesda, Maryland, USA. 6Division of the National Toxicology Program Biomolecular Screening Branch, National Institute of Environmental Health Sciences, Durham, North Carolina, USA. 7Department of Neuropsychiatry, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. 8Laboratoire d’Enzymologie et Biochimie Structurales, Centre National de la Recherche Scientifique, Gif-sur-Yvette, France. 9Department of Physiology, University of Toronto, Ontario, Canada. 10Department of Medicine, University of Toronto, Toronto, Ontario, Canada. 11Neuroscience Department, Centre for Addiction and Mental Health, Toronto, Ontario, Canada. 12Institute of Systems Biology and Bioinformatics, National Central University, Zhongli, Taiwan. 13St. Laurent Institute, Cambridge, Massachusetts, USA. 14Dalhousie School of Public Health, University of Toronto, Toronto, Ontario, Canada. 15These authors contributed equally to this work.

Correspondence should be addressed to A.P. (arturas_petronis@camh.net) or S.K. (klimasau@ibt.lt).

Received 27 April; accepted 2 August; published online 9 September 2012; doi:10.1038/nsmb.2372
from the HpaII-based estimate therefore measures 5-mC. We used biochemical, molecular biological and in silico methods to validate the assay combining glucosylation and restriction-enzyme digestion to measure 5-hmC (Supplementary Figs. 1 and 2 and Supplementary Tables 2 and 3). We verified that the addition of a glucose moiety to 5-hmC confers resistance to HpaII digestion of modified oligonucleotides (Supplementary Figs. 1 and 2). We also confirmed that gDNA modifications can be measured on tiling microarrays (Fig. 1b). Volcano plots demonstrated that nontarget sequences are not affected by DNA glucosylation, as shown by a symmetric zero-centered distribution of intensities, whereas signals from target probes (right) were shifted to the upper right, indicating protection of HpaII targets after DNA glucosylation. These analyses also confirmed that mouse brain samples contain relatively higher 5-hmC densities than do other tissues, as evidenced by a greater shift in the corresponding volcano.

For microarray normalization, we chose an algorithm that corrected for affinity bias due to probe sequence, a known issue for tiling arrays15. We compared two sequence-based normalization algorithms using modification estimates from quantitative polymerase chain reaction (qPCR) (Supplementary Note 1, Supplementary Fig. 2a and Supplementary Tables 2 and 3) and selected an algorithm described previously16 (Supplementary Note 2). We also found that single-probe estimates for 5-hmC or 5-mC provided less bias while maintaining precision for these data, as compared to averaging probe intensities in a local window (Supplementary Table 2). Despite the increased variance in single-probe estimates, biological variability across samples significantly exceeded variability in technical replicates (Supplementary Fig. 2b); for example, probe-wise increase due to biological variability relative to technical variability, mean = 0.52 (arbitrary units), 95% CI, [0.51,0.53]; P > 10^{-18}, one-sample t test). Average probe intensities had the relative magnitude expected from the three treatments:MspI-digested nonglucosylated gDNA had the lowest intensity (for example, in mouse tissues, mean ± s.d., −1.45 ± 1.00 (arbitrary units); 134,521 target probes), followed by MspI-digested glucosylated gDNA (−1.35 ± 1.00). HpaII-digested nonglucosylated gDNA had the highest intensity (−0.81 ± 1.03). Negative values reflect the digestion (under-representation) of target sites relative to the baseline of undigested sequences.

Characterization of 5-mC and 5-hmC in adult-mouse tissues

Using thin-layer chromatography (TLC) and intensities from microarray probes, we verified that 5-hmC levels were the highest in gDNA of the mouse brain, compared to liver, kidney, pancreas and heart (Supplementary Tables 4 and 5). This finding is consistent with previous reports1,5. To investigate the origin of the increased levels of 5-hmC in the brain, we identified genes and intergenic regions with significantly different 5-hmC in the mouse brain compared to other tissues. Of 134,521 probes that overlapped the nonrepetitive resource (Supplementary Table 2), 730 had different 5-hmC in the mouse brain compared to other tissues. Of 4,357 genes tested, 730 had different 5-hmC levels in the brain relative to the other tissues (repeated-measures analysis of variance (ANOVA) P < 10^{-12}, false discovery rate Q < 0.05; Supplementary Table 6). All but one gene had higher 5-hmC and lower 5-mC levels in the brain. Even within the brain, these genes contained above-average 5-hmC levels (P < 10^{-4}, bootstrapping (no replacement); median 5-hmC of randomly sampled genes (R = 10,000)). Separately, we identified 83 differential intergenic probes that also had significantly higher 5-hmC levels in the brain (probe-wise linear regression, total of 60,721 intergenic probes, Q < 0.05; Supplementary Table 7).

We further explored possible associations of genic 5-hmC and steady-state mRNA levels in the mouse data set. We related 5-hmC and 5-mC intensity levels from our microarrays to tissue-specific mRNA levels from a previously published data set17 (Supplementary Note 2). In three out of five tissues investigated, 5-hmC showed a significant increase in genes with higher transcript abundance (Fig. 1c...
Table 1: Statistically overrepresented Gene Ontology categories for genes enriched for 5-hmC in the mouse brain relative to other tissues

| Term                                | Count | %  | P          | Q (BH) |
|-------------------------------------|-------|----|------------|--------|
| GO:0022610, biological adhesion     | 53    | 7.3| 6.3 × 10^{-10} | 1.3 × 10^{-6} |
| GO:0007155, cell adhesion           | 53    | 7.3| 6.3 × 10^{-10} | 1.3 × 10^{-6} |
| GO:0005886, plasma membrane         | 156   | 21.5| 2.4 × 10^{-7} | 8.7 × 10^{-5} |
| GO:0045202, synapse                 | 33    | 4.5| 1.9 × 10^{-6} | 3.5 × 10^{-4} |
| GO:0030054, cell junction           | 41    | 5.6| 1.2 × 10^{-5} | 1.4 × 10^{-3} |
| GO:0044459, plasma membrane part    | 90    | 12.4| 5.6 × 10^{-5} | 5.0 × 10^{-3} |
| GO:0044456, synapse part            | 23    | 3.2| 1.7 × 10^{-4} | 1.2 × 10^{-2} |
| GO:0016337, cell-cell adhesion      | 22    | 3.0| 3.7 × 10^{-5} | 3.9 × 10^{-2} |

and Supplementary Fig. 3). This trend was also observed in the brain, but it did not reach statistical significance (P = 0.17, linear regression; Fig. 1c and Supplementary Fig. 3); the same was true for the 730 genes identified to have higher 5-hmC levels (P = 0.06, bootstrapping randomly-sampled genes (R = 10,000)). In contrast, genic levels of 5-mC significantly decreased for genes with increased mRNA levels in all five tissues (Fig. 1c and Supplementary Fig. 3).

**Synaptic protein genes enriched for 5-hmC in mouse brain**

To determine whether the 730 genes identified above had unifying cellular functions, we performed a functional overrepresentation analysis (DAVID\(^\text{15}\), background set of 4,357 genes tested). We observed that these genes were statistically overrepresented in eight Gene Ontology (GO) terms (total 156 candidate terms), which in turn were associated with synapse function (Table 1). The top overrepresented terms were 'cell adhesion' (Q = 1.3 × 10^{-9}), 'plasma membrane' (Q = 8.7 × 10^{-5}) and 'synapse' (Q = 3.5 × 10^{-4}). These genes also had functional annotation clusters pertaining to ion-channel activity, Rho GTPase signaling and neuronal development (Supplementary Table 8). Notably, some of these 5-hmC-rich genes showed transcript enrichment or specificity for non-neuronal brain cells. We identified the cell-type specificity of the 730 brain enriched genes by using a list from a previously published transcriptomic data set of individual brain-cell populations\(^\text{19}\) (Supplementary Note 2). Using this list, we found that 25 of the brain enriched genes (3%) had enriched transcription in astrocytes, 21 (3%) in oligodendrocytes and 57 (8%) in neurons (Supplementary Table 9). By this definition, a few genes were also specific to the cell type (astrocytes, Bfxf4, Gli3; oligodendrocytes, Elovl7, Cpm; neurons, Nts, Syt1, Nrg3, Trhde, Kcn2, Cstr2).

We tested to see whether the increase in brain 5-hmC observed in the 730 identified genes could be generalized to the entire functional class of synapse-related genes. We compared average 5-hmC in all genes mapped to the synapse-relevant GO terms to that in genes outside these categories\(^\text{20}\) (Supplementary Note 2). Genes within each tested category ('synapse' and 'synapse part', 'cell adhesion', 'plasma membrane') had significantly higher 5-hmC levels (for example, for 'synapse', n = 3,258 probes) than did genes not mapped to the category (n = 70,203; P = 3.2 × 10^{-20}, two-tailed Wilcoxon-Mann-Whitney (WMW) test; Fig. 2a and Supplementary Fig. 4). This effect persisted after controlling for GC content of the probe sequence, a parameter that could artificially influence probe intensity.

**Enrichment of 5-hmC in synapse genes extends to human brain**

We extended the comparison of 5-hmC enrichment in synapse-related genes to the human brain by assaying gDNA from 28 human postmortem brains (frontal cortex Brodmann area 10 (BA10); Stanley Medical Research Institute and Harvard Brain Tissue Resource Center; details in Supplementary Table 10). Human gDNA samples were processed in the same way as for the mouse gDNA samples (Supplementary Note 2) and assayed with Affymetrix 2.0R human whole-genome tiling microarrays E and F (Supplementary Table 1 shows probe counts). Similar to the mouse brain, the human frontal cortex had higher 5-hmC levels within genes that mapped to the GO terms 'synapse', 'synapse part', 'cell adhesion' and 'plasma membrane' ('synapse' and 'synapse part', P = 1.0 × 10^{-4}, one-tailed WMW test; Fig. 2b and Supplementary Fig. 5).

5-hmC marks the exon-intron boundary in the human brain

Several studies performed before the rediscovery of 5-hmC have noted a change in the density of modified cytosines at the exon-intron boundary of genes, with the density being higher on the exonic side\(^\text{21,22}\). These studies, however, used bisulfite sequencing and therefore did not differentiate between 5-mC and 5-hmC. As our restriction enzyme–based assay provided single-nucleotide resolution mapping of DNA modifications, we compared densities of each modification on either side of the exon-intron boundary (linear mixed-effects model, described in Online Methods). Distances to the boundary were measured from the second C of the target sequence (CCGG). DNA-modification differences at exon-intron boundaries are reported for two regions: one immediately adjacent to the boundary (cumulative distance for the first 5 nucleotides, denoted d = 5 base pairs (bp)) and one that captures the general periboundary trend (cumulative distance for the first 20 nucleotides, denoted d = 20 bp). Relevant parameters (for example, probe count) and statistics for all data sets are in Supplementary Table 11. Overall probe intensities for various DNA modifications in all data sets are in Supplementary Table 12.

Levels of all DNA modifications changed at the exon-intron boundary in human frontal cortex samples (n = 28 brain samples; Affymetrix tiling microarrays E and F, covering six chromosomes). We found higher densities of both modifications in exons relative to introns, consistent with previous findings (d = 5 bp, P = 2.8 × 10^{-7}; d = 20 bp, P = 2.8 × 10^{-20}; Supplementary Fig. 5).
In contrast, 5-mC showed no changes closer to the boundary \((d = 5 \text{ bp}, P = 0.57)\) and relatively smaller exonic increases than for 5-hmC for longer periboundary distances \((d = 20 \text{ bp}, P = 9.1 \times 10^{-6}; \text{Fig. 3a,b})\). The change in 5-hmC relative to 5-mC was most evident in the first 5–10 bp from the boundary. For longer distances (up to 50 bp tested), both modifications showed robust increases in exons relative to introns (\(d = 5–50 \text{ bp}\)) (\(\text{Fig. 3b and Supplementary Fig. 6a}\)).

We also mapped DNA modifications in brain samples from patients diagnosed with schizophrenia and bipolar disorder\(^{23}\), collectively termed 'major psychosis' \((n = 54 \text{ brain samples前端 cortex BA10); Affymetrix tiling microarray E})\). Consistent with the findings in the control brain set reported above, gDNA from brains of patients with psychosis also showed a predominant change in 5-hmC at the exon-intron boundary at both distances (\(\text{Fig. 3b, Supplementary Table 11 and Supplementary Fig. 6b}\)).

To ascertain whether the cross-boundary change was unique to brain tissue in humans, we analyzed human liver samples \((n = 13)\) in parallel with age- and sex-matched frontal cortex samples (\(\text{Supplementary Table 10}\)). Here again, brain gDNA samples showed a larger change in 5-hmC than in 5-mC. In contrast, liver gDNA samples showed a predominant change in 5-mC, both at the boundary \((d = 5 \text{ bp}, P = 4.8 \times 10^{-6})\) and at longer distances \((d = 20 \text{ bp}, P = 4.8 \times 10^{-12}; \text{Fig. 3c,d, Supplementary Fig. 6c,d and Supplementary Table 11}\)).

5-mC and 5-hmC boundary changes validated in human brain

We validated the findings in the human frontal cortex using three independent methods. First, to estimate the impact of GC differences in exons and introns, we compared probes representing introns to those representing exons, after matching for GC content and the distance from the boundary (100 iterations to sample a greater number of exonic probes). The overall trends of 5-hmC and 5-mC for matched probes were similar to those for unmatched probes (\(\text{Supplementary Fig. 6e}\)). Cross-boundary differences in 5-hmC were significant \((P < 0.05, \text{geometric mean of 100 iterations})\) for distances from \(d = 5 \text{ bp up to } d = 35 \text{ bp}, \text{whereas 5-mC showed a more modest gradual increase.} \)

Second, we assayed 5-hmC by single-molecule sequencing (SMS), which does not require PCR and microarray hybridization, thus eliminating artifacts due to DNA sequence composition\(^{24}\). Genomic DNA from one frontal cortex sample was digested with MspI with or without prior glucosylation and then subjected to poly(A) end-labeling and sequencing (\(\text{Supplementary Note 2 and Supplementary Table 13}\)). Direct sequencing confirmed a sharp intrinsic decrease in 5-hmC at the exon-intron boundary (one-tailed WMW test, median exonic increase in 5-hmC = 4.6% for \(d = 20 \text{ bp}; P = 9.2 \times 10^{-3}; \text{Fig. 3d and Supplementary Fig. 7}\)). We were unable to accurately measure exonic differences for \(d = 5 \text{ bp}, \text{owing to a small number of reads. However, the exonic increase in 5-hmC is evident for cumulative distances larger than 10 bp (Fig. 3d).} \)

The third approach compared DNA enriched in 5-mC to a fraction depleted in 5-mC (\(\text{Supplementary Note 2}\)). Each fraction was then subjected to the treatment as in \(\text{Figure 1a (n = 6 human frontal cortex samples; Supplementary Fig. 8a).\) \text{We expected that 5-mC levels in the brain actually changed across the exon-intron boundary, we would detect a greater cross-boundary change in the 5-mC-rich fraction relative to the depleted fraction. We did not observe such a change at the exon-intron boundary; in fact, intrinsic levels of 5-mC were slightly higher (P = 0.05, d = 20 bp).} \)

The above three control experiments collectively show that in the human frontal cortex, the change in 5-hmC at the exon-intron boundary exceeds that of 5-mC.

DNA modification at exon-intron boundaries in mouse tissues

To determine whether the change in 5-hmC at exon-intron boundaries is unique to the human frontal cortex, we examined DNA in multiple mouse organs including the mouse brain. Brain samples (from 8-week-old male C57/BL6 mice) were split into frontal cortex and the remainder (including cerebellum and brain stem) \((n = 15 \text{ samples per group})\) and analyzed with Affymetrix mouse tiling microarray A. The frontal cortex was analyzed separately to match the brain region investigated in human samples. Consistent with patterns in human brain, at \(d = 20 \text{ bp from the exon-intron boundary, the changes in levels of DNA modifications were mainly due to 5-hmC both in frontal cortex (P = 7.5 \times 10^{-6}, \text{linear mixed-effects model}) and in the rest of the brain (P = 3.0 \times 10^{-8}; \text{Fig. 4a,b and Supplementary Table 11})\). In these contexts, the change in 5-hmC was not evident at the immediate boundary (for example, \(P = 0.35\) at \(d = 5 \text{ bp in the frontal cortex}) but rather was gradual over a region adjacent to the boundary. In contrast, 5-mC exhibited a more dramatic immediate cross-boundary change \((P = 9 \times 10^{-4} \text{ at } d = 5 \text{ bp})\) (\(\text{Fig. 4a,b}\)).
For non-neural organs, we used the mouse data set described previously in this work (Supplementary Table 5; 36 samples from liver, pancreas, kidney and heart; microarrays A and G). In agreement with the markings seen in the human liver, mouse organs of non-neural origin primarily showed changes in cross-boundary densities of 5-mC ($d = 5 \text{ bp}$, $P = 0.03$), with a nonsignificant change in 5-hmC ($d = 5 \text{ bp}$, $P = 0.25$) (Fig. 4c and Supplementary Table 11). Although both types of DNA modifications show relatively higher exonic densities at the larger periboundary regions ($d = 20 \text{ bp}$), the magnitude of change in 5-mC ($P = 3.4 \times 10^{-19}$) exceeded that of 5-hmC ($P = 5.7 \times 10^{-5}$) (Fig. 4c).

5-mC and 5-hmC at the exon-intron boundary in cell lines

We next assayed a mouse neuronal cell line, which had undetectable (<1%) global levels of 5-hmC measured by TLC (mHypoA-2/24; $n = 18$ samples, Supplementary Table 12). The microarray analysis, which is more sensitive to region-specific differences than is TLC, detected a marginally significant change in 5-mC at $d = 20 \text{ bp}$ ($P = 0.045$, linear mixed-effects model) and a highly significant difference in 5-mC ($P = 1.6 \times 10^{-17}$) (Supplementary Fig. 8b and Supplementary Table 11).

We also examined boundary differences in a B-lymphocyte data set in which cells were treated with suberoylanilide hydroxamic acid (SAHA). Histone deacetylase (HDAC) inhibitors such as SAHA may promote DNA demethylation and have also been shown to induce changes in pre-mRNA splicing. We examined whether SAHA induces changes in DNA modifications at the exon-intron junction.

Vehicle-treated (no SAHA) B lymphocytes showed no 5-hmC exon-intron differences at $d = 20 \text{ bp}$ ($P = 0.27$); however, SAHA-treated cells showed an increase in the cross-boundary differences in a manner roughly corresponding to increasing doses of SAHA ($P = 0.01$, $8.8 \times 10^{-4}$ and $3.6 \times 10^{-4}$ for 0.01 μM, 0.02 μM and 0.1 μM of SAHA, respectively) (Supplementary Fig. 9 and Supplementary Table 11).

Exonic 5-hmC and exon inclusion levels

Changes in DNA modifications at exon-intron boundaries may affect exon recognition and exon inclusion levels. To investigate this possibility, we measured 5-hmC as a function of exon inclusion levels, using RNA sequencing data for human frontal cortex and liver (described in Online Methods). Exon inclusion levels were measured as the proportion of transcripts that included a given exon; exons were classified as being alternatively spliced (exon included in ≤80% of gene transcripts) or constitutive (exon included in >100% of gene transcripts) (Online Methods). We observed that at $d = 20 \text{ bp}$ the median 5-hmC density was lower in alternatively spliced exons relative to constitutive exons ($P = 0.04$, two-tailed WMW test; Supplementary Table 14 shows probe counts and exon sampling), but it was not different for 5-mC ($d = 20 \text{ bp}$, $P = 0.61$). The number of probes at $d = 5 \text{ bp}$ was too small for a meaningful comparison. The 5-hmC effect was even stronger at the level of the whole exon ($P = 8.4 \times 10^{-5}$); here a marginal change was observed in 5-mC levels ($P = 0.025$, Fig. 5a). The increase in 5-hmC was still nearly significant after probes in alternative and constitutive exons were matched for GC content, but in this case the contrast with 5-mC was less pronounced (1,000 iterations, 197 probes per group; geometric mean, $P = 0.07$ for 5-hmC and $P = 0.08$ for 5-mC). In the liver, neither 5-hmC nor 5-mC was different among the two types of exons ($P = 0.13$ for 5-hmC, $P = 0.38$ for 5-mC, Supplementary Table 14 and Fig. 5b).

We found separately that average exonic 5-hmC in intronless or single-exon genes was lower than in genes with multiple exons. For this analysis, we used tiling-array data for control human brains over six chromosomes (arrays E and F, 28 samples, sample median used for each probe; RefSeq gene definitions were used; within each category, duplicate bases were collapsed). We found that exonic 5-hmC in single-exon genes was statistically different from that in multi-exon genes (median, intronless = 0.06 (arbitrary units), multi-exon = 0.073; $P = 0.026$, two-tailed WMW test; $n$, intronless = 1,256 probes, multi-exon = 22,500 probes). This finding is consistent with an exonic change in 5-hmC due to gene splicing.

Collectively, the results suggest that the density of 5-hmC proximal to splice sites and within exons could affect splicing and exon inclusion levels in the mammalian brain.

---

**Figure 4** 5-hmC marks exon-intron boundaries in the mouse brain but not in mouse organs of non-neural origin. (a, b) Periboundary modifications in DNA in the mouse frontal cortex (a) and remainder (b) ($n = 15$ animals per group, 3 chromosomes). Graphical depiction and statistical details are as for Figure 3. Left and middle, median intensities around the exon-intron boundary; right, relative exonic increase at various cumulative distances from the boundary. (c) DNA modifications in non-neural mouse organs ($n = 24$ animals, 6 chromosomes; DNA from heart, liver, pancreas and kidney).

**Figure 5** Constitutive exons have higher 5-hmC levels than do alternatively-spliced exons in human brain. (a, b) Comparison of probe-level intensities of each modification in alternatively-spliced (Alt) and constitutive (Const) exons for human brain (a) and human liver (b). $P$ values are from two-tailed WMW test ($α = 0.05$). Probe count shown under box plots for each panel; other details in Supplementary Table 14.
DISCUSSION

In this study, we adapted a 5-hmC detection strategy that uses glucosylation-induced resistance to restriction enzymes for tiling microarray-based mapping of 5-hmC. The glucosylation-based detection of 5-hmC has been successfully used in other studies. Our microarray-based quantification of DNA modifications (verified by TLC) showed significantly higher levels of 5-hmC in brain as compared to other tissues, which is consistent with previous reports. We identified a large number of genes with higher densities of 5-hmC in the brain compared to those in heart, liver, kidney and pancreas. More generally, we discovered a trend whereby 5-hmC in the gene body increases with increased transcription of the corresponding gene, whereas 5-mC decreases. This is consistent with the observation that 5-hmC prevents the binding of transcriptional repressor proteins and is found within actively transcribed genes. Notably, the association between genomic 5-hmC and mRNA levels was weakest in the brain (P = 0.17), suggesting that 5-hmC may have other functions in brain cells, in addition to the regulation of gene activity.

Functional annotation analysis revealed a statistical overrepresentation of terms pertaining to synaptic plasticity in genes enriched for 5-hmC in the brain. In particular, annotation clusters were composed of protein groups involved in distinct aspects of synaptic remodeling: ion channels, members of the Rho GTPase signaling pathway and axon-guidance molecules. Notably, most genes in these clusters encode proteins that are functionally located at the plasma membrane rather than being cytosolic. The observation that 5-hmC is overrepresented in the genes controlling synaptic plasticity may shed a new light on the epigenetics of learning and memory. Adult animals with conditional double knockouts of genes encoding DNA methyltransferases 1 and 3b (Dnmt1 and Dnmt3b) in the cerebral cortex showed learning and memory defects in hippocampal-dependent learning tasks, which suggests that changes in DNA methylation occur in post-mitotic neurons. Fear conditioning, an experimental paradigm for emotional learning, results in the demethylation and transcriptional activation of reelin (RELN). Recent mappings of 5-hmC in the brain have demonstrated generation of 5-hmC and demethylation in response to neuronal activity, but the degree to which the two correspond is not completely clear. In the absence of DNA replication in postmitotic neurons, it is possible that loci undergoing gene reactivation by DNA demethylation accumulate 5-hmC over time. Experiments that map 5-hmC over the course of multiple gene reactivations will explore whether, in the context of synaptic activity, this base marks a stable epigenetic state or is an intermediate in cycles including complete demethylation (that is, conversion to unmodified C).

It is possible that 5-hmC and 5-mC may affect the process of pre-mRNA splicing in addition to transcriptional regulation. Our finding of boundary changes in the examined cytosines replicates earlier observations that used bisulfite sequencing. The separation of the two DNA modifications in our study showed that 5-hmC, rather than 5-mC, accounts for most of the density difference at the immediate exon-intron boundary in human frontal cortex samples. This finding was validated by three control experiments: comparison of GC-matched probes, SMS and methyl binding domain–enriched 5-mC mapping. Whereas the mouse brain samples showed a predominant change in 5-hmC at longer periboundary distances, the main change within the first 10 bp of the boundary was in 5-mC.

The non-neural tissues investigated included human liver and four mouse organs (liver, pancreas, heart and kidney). Whereas these tissues had levels of genomic 5-hmC density measurable by TLC, modification changes at the exon-intron boundary were mainly due to 5-mC. In contrast to the pattern observed in the brain, periboundary changes in 5-hmC were relatively minor in these tissues. These findings suggest that any splicing-related functions that DNA modification may have in non-neural organs are mediated mainly by 5-mC rather than 5-hmC. Separately, we found that B lymphocytes show increased 5-hmC differences at the exon-intron boundary upon treatment with SAHA. SAHA is a chemotherapeutic agent that acts as an HDAC inhibitor and may promote DNA demethylation. As the SAHA dose used in this study is comparable to the plasma concentrations in patients treated for cancer with SAHA, 5-hmC changes at the exon-intron boundary may also be expected to take place in vivo. Hence, it is also possible that such changes in 5-hmC in response to HDAC inhibition could also contribute to alternative splicing.

Previous studies have reported that exons are enriched in histone H3 trimethylation at Lys36 and Lys4 and dimethylation at Lys27, relative to flanking intronic regions. These modifications can recruit splicing regulators to affect alternative splicing of nascent transcripts. Whereas the resolution of histone maps is limited by the size of nucleosomal DNA (147 bp), our DNA modification studies have identified changes in modification levels within 20 bp (and in some cases within 5 bp) surrounding the exon-intron boundary. This precision argues for a specific effect at the exon-intron boundary as well as a possible difference between exons and introns as a whole, as described in earlier studies. Our finding that 5-hmC densities are lower in alternatively spliced exons relative to constitutive exons complements the observation that trimethylated histone H3 Lys36 is less enriched in alternatively spliced exons. DNA methylation has been shown to modulate exon inclusion levels by influencing the rate of transcript elongation in B lymphocytes. Lack of DNA methylation can promote exon inclusion by causing ‘pauses’ in RNA polymerase II–mediated elongation, and it may also affect RNA polymerase II elongation-dependent changes in alternative splicing by affecting the binding of the transcriptional repressor CTCF. There may also be other mechanisms by which DNA methylation could influence exon inclusion levels, for example, through the recruitment of splicing factors by methyl-binding proteins.

Our findings suggest that tissue-specific distributions of 5-hmC or 5-mC at the exon-intron boundary and within genes may simultaneously influence both transcription and splicing. The direction of the influence remains unclear, as transcription and splicing may also affect epigenetic DNA modifications, and mechanisms for cross-talk likely exist between epigenetic regulation, splicing, and transcription.

METHODS

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

Accession codes. Data have been deposited in GEO, with accession numbers GSE40167 (for the SuperSeries), GSE40166 and GSE40158.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank N. Miller and C. Austin (National Center for Advancing Translational Sciences, US National Institutes of Health, Bethesda, Maryland, USA) for providing cell lines for the SAHA experiment. We also thank J.-S. Doucet (Centre for Addiction and Mental Health Toronto, Canada) for assistance with mouse brain sample preparation. This research has been supported by the Canadian Institutes of Health Research (grants 199170 and 186007) and the US National Institutes of Health (grants MH074127, MH088413, DP3DK085698, HG005758 and HG004353) to A.P. and Canadian Institutes of Health Research grants to B.J.B., by the Research Council of Lithuania and the French Ministry of Foreign and European affairs under the Lithuania-France bilateral collaboration program Gilbert (grant TAP-13/2011–2012) to S.M. and S.K., and by the European Social Fund under the Global Grant measure (grant VPI-3.1-ŠMM-07-K-01-105) to S.K. A.P. is supported as a Tapscott Chair in Schizophrenia Studies at the University.
of Toronto and by the Ontario Mental Health Foundation. M.I. is supported by a Human Frontier Science Program Organization Long Term Fellowship.

AUTHOR CONTRIBUTIONS

S.K., A.P.T.K. and S.P. conceived of the study and designed the experiments. S.M., E.K., Z.L. and S.K. developed and verified the methods for 5-hmC detection. T.K., M.P., M.T., P.J., C.P., V.L., A.N., D.B. and A.H.C.W. conducted the microarray experiments. S.P., K.K., S.C.W. and R.K. developed bioinformatics methods and analyzed microarray data. P.K., T.K., S.P. and A.P. contributed to the design and data analysis of the Helicos sequencing experiment. M.X. and R.T. generated SAHA-treated cell lines. M.I. and B.J.B. generated the lists of alternative and constitutive exons and performed 5-hmC comparisons. T.K., S.P., V.L., S.K. and A.P. wrote the manuscript. All authors reviewed the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nsmb.2372
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324, 929–930 (2009).
2. Penn, N.W., Suwalski, R., O’Riley, C., Bojanowski, K. & Yura, R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. Biochem. J. 126, 781–790 (1972).
3. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 900–905 (2009).
4. Huang, Y. et al. The behaviour of 5-hydroxymethylcytosine in bisulphite sequencing. PLoS ONE 5, e8888 (2010).
5. Glocisch, D. et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS ONE 5, e15367 (2010).
6. Deaton, A.M. & Bird, A. CpG islands and the regulation of transcription. Genes Dev. 25, 1010–1022 (2011).
7. Jin, S.-G., Wu, X., Li, A.X. & Pfeifer, G.P. Genomic mapping of 5-hydroxymethylcytosine in the human brain. Nucleic Acids Res. 39, 5015–5024 (2011).
8. Song, C.-X. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat. Biotechnol. 29, 68–72 (2011).
9. Guo, J.J., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxilation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell 145, 423–434 (2011).
10. He, Y.F. et al. Tet-mediated formation of 5-carboxy-cytosine and its excision by TDG in mammalian DNA. Science 333, 1303–1307 (2011).
11. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333, 1303–1307 (2011).
12. Zhang, L. et al. Thymine DNA glycosylase specifically recognizes 5-carboxycytosine-modified DNA. Nat. Chem. Biol. 8, 328–330 (2012).
13. Huang, L.H., Farnet, C.M., Ehrlich, K.C. & Ehrlich, M. Digestion of highly modified bacteriophage DNA by restriction endonucleases. Nucleic Acids Res. 10, 1579–1591 (1982).
14. Josse, J. & Kornberg, A. Glucosylation of deoxyribonucleic acid. III. α- and β-glucosyl transferases from T4-infected Escherichia coli. J. Biol. Chem. 237, 1968–1976 (1962).
15. Liu, X.S. Getting started in tiling microarray analysis. PLOS Comput. Biol. 3, 1842–1844 (2007).
16. Potter, D.P., Yan, P., Huang, T.H.M. & Lin, S. Probe signal correction for differential methylation hybridization experiments. BMC Bioinformatics 9, 453 (2008).
17. Lattin, J.E. et al. Expression analysis of G protein-coupled receptors in mouse macrophages. Immuneine Res. 4, 5 (2008).
18. 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).
19. Cahoy, J.D. et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J. Neurosci. 28, 264–278 (2008).
20. Balsa-Canto, E. & Banga, J.R. AMIGO, a toolbox for advanced model identification in systems biology using global optimization. Bioinformatics 27, 2311–2313 (2011).
21. Chodavarapu, R.K. et al. Relationship between nucleosome positioning and DNA methylation. Nature 466, 410–416 (2010).
22. Laurent, L. et al. Dynamic changes in the human methylome during differentiation. Genome Res. 20, 320–331 (2010).
23. Torrey, E.F., Webster, M., Knable, M., Johnston, N. & Yolken, R.H. The Stanley Foundation brain collection and neuropathology consortium. Schizophr. Res. 44, 151–155 (2000).
24. Hart, C. et al. Single-molecule sequencing: sequence methods to enable accurate quantitation. Methods Enzymol. 472, 407–430 (2010).
25. Ou, J.N. et al. Histone deacetylase inhibitor trichostatin A induces global and gene-specific DNA demethylation in human cancer cell lines. Biochem. Pharmacol. 73, 1297–1307 (2007).
26. Hnilicova, J. et al. Histone deacetylation activity modulates alternative splicing. PLoS ONE 6, e16727 (2011).
27. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. Nature 478, 343–348 (2011).
28. Davis, T. & Vassila, R. High sensitivity 5-hydroxymethylcytosine detection in human tissue. J. Vis. Exp. 53, 857–869 (2007).
29. Jin, S.-G., Kadam, S. & Pfeifer, G.P. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res. 38, e125 (2010).
30. Kondo, E., Gu, Z., Horii, A. & Fukushima, S. The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. Mol. Cell Biol. 25, 4388–4396 (2005).
31. Vlaminck, V. et al. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Mol. Cell 29, 929–939 (2008).
ONLINE METHODS

Experimental outline. Our experimental strategy exploited the ability of the BGT enzyme of T4 phage to transfer a monomeric glucose residue from uridine diphospho-α-D-glucose (UDP-Glc) to the exocyclic hydroxyl group of 5-hmC in double-stranded DNA. This specific addition of a glucose moiety to 5-hmC blocks access of methylation-insensitive endonucleases to their target sites, thus making the DNA fragment resistant to digestion.13,14 Here we used the isochizomer restriction enzymes HpaII and MspI. HpaII cleaves DNA only when the CpG dinucleotide is unmethylated at CCGG, whereas MspI cleaves the CCGG sites independently of any natural CpG modification, including 5-mC and 5-hmC.47,48

Microarray experiments were performed as illustrated in Figure 1. Each genomic DNA sample was split into three parts, one of which was subjected to BGT glucosylation and two of which were not glucosylated (details in Supplementary Note 2). One unglucosylated part was treated with MspI, serving as a baseline for digestion. The second unglucosylated part was treated with HpaII, which spares any modified cytosines at the recognition site. The BGT-glucosylated part was treated with MspI, and 5-hmC was measured as the decrease in digestion of this part, relative to the unglucosylated part. The difference between HpaII and MspI restriction efficacy on unglucosylated DNA reflected 5-hmC and 5-mC combined, and the difference between the former measurement and 5-hmC was the measure of 5-mC.

Sample description. This study used data sets from various mouse organs, human brain and liver and two cell lines. Male C57BL/6j mouse brain tissues (frontal cortex and remaining brain) and other organs (liver, pancreas, kidney and heart) were obtained from 8-week-old mice (Supplementary Table 5). All human brain samples were obtained from the Stanley Medical Research Institute brain collection (courtesy of M.B. Knable, E. Fuller Torrey, M.J. Webster and R.H. Yolken, Chevy Chase, Maryland, USA) and from the Harvard Brain Tissue Resource Center (courtesy of F. Benes, Belmont, Massachusetts, USA). Human liver samples were obtained from commercial databanks (Curline (USA) and Cambridge Bioscience (UK)). Liver samples were matched by age and sex to brain samples. Further details are provided in Supplementary Table 10. Genomic DNA isolation was performed by standard proteinase K digestion and phenol-chloroform extraction.

Sample preparation for glucosylation and restriction digestion. From each tissue, gDNA was sheared to 200-bp fragment length by using a Covaris S2 sonifier (Covaris, USA). Sheared DNA was end-filled in the presence of T4 DNA polymerase (5 U) at 11 °C for 20 min and purified using PEG buffer from the QIAquick Nucleotide Removal Kit (Qiagen). The purified 1 μg of blunt-ended DNA was ligated with Alu-25 adaptors (10 μM) (Supplementary Table 15) in the presence of T4 DNA ligase (10 U) at 22 °C for 3 h, inactivated at 65 °C for 10 min and purified with the QIAquick Nucleotide Removal Kit. One-third (~300 ng) of the purified DNA was treated with 200 μM uridine-5′-diphospho-α-D-glucose (UDP-Glc, Sigma), 80 ng BGT in 100 mM Tris-HCl, pH 8.0, and 25 mM MgCl2 for 3 h at 37 °C and purified again. Thereafter, 200 ng of glucosylated DNA and 200 ng of untreated DNA were subjected to MspI digestion in a final volume of 20 μl at 37 °C for 1 h, followed by enzyme inactivation at 80 °C for 10 min. The same protocol was performed on 200 ng of untreated DNA using HpaII. Following the observation that some of our data points did not follow the expected normal distribution, we used the log difference between glucosylated DNA and native DNA, following treatment with MspI (Fig. 1a). The sum of all the DNA modifications was measured as the log difference between HpaII-treated glucosylated DNA and MspI-treated unglucosylated DNA. 5-mC was estimated by the difference between HpaII and MspI restriction efficacy on unglucosylated DNA.

Following normalization, probes overlapping or collinear regions were excluded (RepeatMasker, simple repeats and segmental duplications; genomic annotations from UCSC genome browser, build mm8 for mouse, hg18 for human). Information on arrays and probe counts is listed in Supplementary Table 1. Unless otherwise specified, analysis was done using R software and BioConductor.35

For the exon-intron boundary analysis, probes were required to have a CCGG in both the Affymetrix probe sequence and the chromosomal sequence downloaded from UCSC; probes that did not meet this criterion were excluded.

Identification of 5-hmC enriched regions in mouse brain. Probes were filtered for those overlapping genes on either strand (RefSeq genes (RefGene) downloaded from UCSC, mm8; range between txStart and txEnd columns). Probes overlapping multiple genes (as defined by MGI gene symbol) were excluded, as were genes containing exactly one probe, resulting in 73,461 probes over 4,357 genes (range, 2–238 probes per gene; mean, 17). A repeated-measures ANOVA was conducted on probes in each gene, with tissue as a between-groups factor (‘brain’, 11 samples; ‘other’, 36 samples; liver, heart, kidney, pancreas, 9 samples each) and the biological sample as a within-groups error term. Probes within a gene were neither averaged across samples nor collapsed within a gene. Gene-wise nominal P values were adjusted using the Benjamini-Hochberg method of false-discovery-rate control (BH; α = 0.01). Of genes with Q < 5% (n = 730), all but one had higher 5-hmC in the brain, and these are therefore called ‘brain enriched’. Calculation of exon-intron boundary differential. Boundary regions were defined as regions within a certain distance on both sides of an exon start or end (RefSeq genes, internal exons only). Where multiple exons had the same genomic extent (that is, location on chromosome, start and end), only one such exon was retained. For each probe, the distance from the modifiable second C of CCGG was computed to the nearest exon (or intron) boundary; this value is the boundary distance of the probe (a value of 0 was exonic in these calculations). Statistics were computed on probes that lay ≤5 or ≤20 bp on either side of the junction. The result at a distance of 20 bp reflected the general trend observed at the boundary, whereas that at 5 bp reflected immediate cross-boundary change. A linear mixed-effects model was used to test probe intensity differences between the exonic and intronic side of the junction. Junction side was used as the fixed-effects term, and array type and sample were treated as random-effects terms (Supplementary Note 2). α = 0.025 was used to indicate a statistically significant result.

Correlation of 5-hmC with exon inclusion levels. RNA-seq results from liver and cortex were analyzed by using R (RefSeq genes, internal exons only). Where multiple exons had the same genomic extent (that is, location on chromosome, start and end), only one such exon was retained. For each probe, the distance from the modifiable second C of CCGG was computed to the nearest exon (or intron) boundary; this value is the boundary distance of the probe (a value of 0 was exonic in these calculations). Statistics were computed on probes that lay ≤5 or ≤20 bp on either side of the junction. The result at a distance of 20 bp reflected the general trend observed at the boundary, whereas that at 5 bp reflected immediate cross-boundary change. A linear mixed-effects model was used to test probe intensity differences between the exonic and intronic side of the junction. Junction side was used as the fixed-effects term, and array type and sample were treated as random-effects terms (Supplementary Note 2). α = 0.025 was used to indicate a statistically significant result.

Microarray data preprocessing. We distinguished between three types of probes on the whole-genome tiling microarray: ‘target probes’, which contain the recognition sequence, ‘non-target probes’, which do not contain the recognition sequence but could lie up to 200 bp upstream or downstream of a target probe (sheared DNA fragments were, on average, 200 bp) and ‘non-target probes’ that neither contain a target sequence nor are within 200 bp (on either side) of the target sequence. Nontarget probes are unaffected by enzymatic cleavage and therefore can be used as a baseline for array normalization. After comparing different array preprocessing algorithms (Supplementary Note 1), we chose to use a sequence-based algorithm after a previous study.16

The model (Supplementary Note 2, equation (1)) was applied to nontarget probes matched proportionally to target probes by GC content. Normalized intensities were obtained by subtracting the fitted value from raw probe intensities, resulting in a normal distribution of probe-level intensities. Each chip was independently normalized. All downstream analyses were carried out at the single-probe level (that is, without windowing or peak-calling) and solely on target probes (henceforth referred to as ‘probes’). 5-hmC was measured as the log difference between glucosylated DNA and native DNA, following treatment with MspI (Fig. 1a). The sum of all the DNA modifications was measured as the log difference between HpaII-treated glucosylated DNA and MspI-treated unglucosylated DNA. 5-mC was estimated by the difference between all DNA modifications and 5-hmC.

Correlation of 5-hmC with exon inclusion levels. RNA-seq results from liver and cortex were analyzed by using R (RefSeq genes, internal exons only). Where multiple exons had the same genomic extent (that is, location on chromosome, start and end), only one such exon was retained. For each probe, the distance from the modifiable second C of CCGG was computed to the nearest exon (or intron) boundary; this value is the boundary distance of the probe (a value of 0 was exonic in these calculations). Statistics were computed on probes that lay ≤5 or ≤20 bp on either side of the junction. The result at a distance of 20 bp reflected the general trend observed at the boundary, whereas that at 5 bp reflected immediate cross-boundary change. A linear mixed-effects model was used to test probe intensity differences between the exonic and intronic side of the junction. Junction side was used as the fixed-effects term, and array type and sample were treated as random-effects terms (Supplementary Note 2). α = 0.025 was used to indicate a statistically significant result.

Correlation of 5-hmC with exon inclusion levels. RNA-seq results from liver and cortex were analyzed by using R (RefSeq genes, internal exons only). Where multiple exons had the same genomic extent (that is, location on chromosome, start and end), only one such exon was retained. For each probe, the distance from the modifiable second C of CCGG was computed to the nearest exon (or intron) boundary; this value is the boundary distance of the probe (a value of 0 was exonic in these calculations). Statistics were computed on probes that lay ≤5 or ≤20 bp on either side of the junction. The result at a distance of 20 bp reflected the general trend observed at the boundary, whereas that at 5 bp reflected immediate cross-boundary change. A linear mixed-effects model was used to test probe intensity differences between the exonic and intronic side of the junction. Junction side was used as the fixed-effects term, and array type and sample were treated as random-effects terms (Supplementary Note 2). α = 0.025 was used to indicate a statistically significant result.

% exon inclusion = \( \frac{100 \times (\sum(C_{G1}) + \sum(AG_{C2})))}{(\sum(C_{G1}) + \sum(AG_{C2}) + \sum(C_{G2}) + \sum(C_{C1}))} \)

where \( C_{G} \) is any possible splicing donor upstream from the alternative exon (including \( C_{1} \)), and \( C_{C} \) is any possible splicing acceptor downstream from the alternative exon (including \( C_{2} \)). Exons with multiple acceptor or donor splice
sites and those not supported by RNA-seq were excluded. We defined an exon as alternatively spliced if the inclusion level was less than 80% and as constitutive if inclusion was 100%. DNA methylation was assessed in those exons that overlapped 'target probes' (probes containing CCGG) (probe counts are shown in Supplementary Table 14). Methylation levels of exons were computed at the whole-exon level and at \( d = 20 \) bp away from the boundary. Sample-averaged median probe intensities were compared between probes overlapping constitutive and alternative exons (two-tailed WMW test, \( \alpha = 0.05 \)).

47. Jin, S.-G., Kadam, S. & Pfeifer, G.P. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.* **38**, e129 (2010).

48. Tardy-Planechaud, S., Fujimoto, J., Lin, S.S. & Sowers, L.C. Solid phase synthesis and restriction endonuclease cleavage of oligodeoxynucleotides containing 5-(hydroxymethyl)-cytosine. *Nucleic Acids Res.* **25**, 553–559 (1997).

49. Larivière, L., Kurzeck, J., Aschke-Sonnenborn, U., Ruger, W. & Morera, S. Crystallization and preliminary crystallographic study of a ternary complex between the T4 phage beta-glucosyltransferase, uridine diphosphoglucose and a DNA fragment containing an abasic site. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1484–1486 (2002).

50. Ichiyanagi, K. Inhibition of MspI cleavage activity by hydroxymethylation of the CpG site: a concern for DNA modification studies using restriction endonucleases. *Epigenetics* **7**, 131–136 (2012).

51. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

52. Vittinghoff, E., Glidden, D.V., Shiboski, S.C. & McCulloch, C.E. Repeated Measures Analysis, in *Regression Methods in Biostatistics* Ch. 8, 253–290 (Springer, 2005).