

# Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage

Mirang Kim<sup>1,†</sup>, Young-Kyu Park<sup>2,†</sup>, Tae-Wook Kang<sup>1,‡</sup>, Sang-Hun Lee<sup>4</sup>, Yong-Hee Rhee<sup>4</sup>, Jong-Lyul Park<sup>1,5</sup>, Hee-Jin Kim<sup>1,5</sup>, Daeyoung Lee<sup>3</sup>, Doheon Lee<sup>2</sup>, Seon-Young Kim<sup>1,5,\*</sup> and Yong Sung Kim<sup>1,5,\*</sup>

<sup>1</sup>Medical Genomics Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305-806, Korea <sup>2</sup>Department of Bio and Brain Engineering and <sup>3</sup>Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea <sup>4</sup>Department of Biochemistry and Molecular Biology, College of Medicine, Hanyang University, Seoul 133-791, Korea and <sup>5</sup>Department of Functional Genomics, University of Science and Technology (UST), Daejeon 305-350, Korea

Received March 12, 2013; Revised August 26, 2013; Accepted September 13, 2013

DNA methylation and hydroxymethylation have been implicated in normal development and differentiation, but our knowledge is limited about the genome-wide distribution of 5-methylcytosine (5 mC) and 5-hydroxymethylcytosine (5 hmC) during cellular differentiation. Using an *in vitro* model system of gradual differentiation of human embryonic stem (hES) cells into ventral midbrain-type neural precursor cells and terminally into dopamine neurons, we observed dramatic genome-wide changes in 5 mC and 5 hmC patterns during lineage commitment. The 5 hmC pattern was dynamic in promoters, exons and enhancers. DNA hydroxymethylation within the gene body was associated with gene activation. The neurogenesis-related genes *NOTCH1*, *RGMA* and *AKT1* acquired 5 hmC in the gene body and were up-regulated during differentiation. DNA methylation in the promoter was associated with gene repression. The pluripotency-related genes *POU5F1*, *ZFP42* and *HMGA1* acquired 5 mC in their promoters and were down-regulated during differentiation. Promoter methylation also acted as a locking mechanism to maintain gene silencing. The mesoderm development-related genes *NKX2-8*, *TNFSF11* and *NFATC1* acquired promoter methylation during neural differentiation even though they were already silenced in hES cells. Our findings will help elucidate the molecular mechanisms underlying lineage-specific differentiation of pluripotent stem cells during human embryonic development.

## INTRODUCTION

DNA methylation is an essential epigenetic mechanism for controlling normal development and in some cases promoting disease progression (1). Methylation at the 5 position of cytosine (5 mC) is catalyzed by DNA methyltransferases (DNMTs) such as DNMT1, DNMT3A and DNMT3B (2). DNA methylation plays an important role in silencing tissue-specific genes, imprinted genes and repetitive elements (3,4). Although methylation is a relatively stable epigenetic control mechanism for long-term silencing, dynamic changes

in methylation patterns occur during development and differentiation (5).

Recently, 5-hydroxymethylcytosine (5hmC) was observed in the DNA of neural tissue and embryonic stem (ES) cells (6,7). DNA hydroxymethylation is mediated by the ten-eleven translocation (TET)-family enzymes such as TET1, TET2 and TET3 (8). Although 5 hmC is an intermediate during passive and active demethylation (9,10), the relatively high levels of 5 hmC in ES cells and some differentiated cells suggest that 5 hmC plays an important role in gene regulation that is separate from the role of 5 mC (7,11).

\*To whom correspondence should be addressed at: Medical Genomics Research Center, Korea Research Institute of Bioscience and Biotechnology, 111 Gwahangno, Yuseong-gu, Daejeon 305–806, Korea. Tel: +82 428798116 (S.-Y.K.)/+82 428798110 (Y.S.K.); Fax: +82 428798119 (S.-Y.K.)/+82 428798119 (Y.S.K.); Email: kimsy@kribb.re.kr (S.-Y.K.)/yongsung@kribb.re.kr (Y.S.K.)

<sup>†</sup>M.K. and Y.-K.P. contributed equally to this work.

<sup>‡</sup>Present address: Bioinformatics Division, Bio Medical Laboratories (BML), Daejeon 305-301, Korea.

Human ES (hES) cells are the only available *in vitro* model system for studying lineage commitment and cell fate decisions in humans (12). Embryonic midbrain-type neural precursor (NP) cells derived from hES cells proliferate *in vitro* and can differentiate into dopamine (DA) neurons (13,14). Using this *in vitro* model system, we explored epigenetic changes in DNA methylation during neural lineage commitment of pluripotent stem cells. We used two techniques for genome-wide analysis of 5 mC/5 hmC. The first, MBD-seq, involves high-throughput sequencing of methylated DNA fragments captured by methyl-CpG-binding domain protein 2 (15). The second, hMeDIP-seq, uses 5 hmC-specific antibodies for genome-wide analysis of 5 hmC patterns (16). We also used bisulfite sequencing (BS) and Tet-assisted BS (TAB-Seq) (11) to confirm 5 mC and 5 hmC changes at specific loci to single-base resolution. These methods revealed dynamic changes in the patterns of 5 mC/5 hmC during neural differentiation of hES cells.

We further investigated the role of DNA hydroxymethylation and methylation in regulating gene expression by examining global patterns within gene bodies and promoters and by examining local patterns within the promoter regions of neural, pluripotent- and mesoderm development-related genes during neural differentiation.

## RESULTS

### Profiling of the DNA methylome, hydroxymethylome and transcriptome during neural differentiation of hES cells

To study the extent and role of DNA methylation and hydroxymethylation during lineage commitment, H9 hES cells were guided to differentiate into ventral midbrain-type NP cells *in vitro*. Approximately 92% of cells were positive for the NP-specific intermediate filament protein NESTIN at this stage (Fig. 1A; (17)). The NP cells were then terminally differentiated into DA neurons. Approximately 43% of these cells were positive for the DA neuron-specific marker tyrosine hydroxylase by immunofluorescence (Fig. 1A; (17)). We performed MBD-seq to characterize the DNA methylome of the three differentiation states (hES cells, NP cells and DA neurons) and hMeDIP-seq to characterize the DNA hydroxymethylome. We also performed expression array analysis using a human whole-genome expression BeadChip to profile the transcriptome of the three states.

The expression array analysis showed that the ES cell markers *POU5F1* (alias *OCT4*) and *NANOG* were down-regulated in NP cells but that the gene for *SOX2*, which inhibits mesendodermal differentiation and promotes neural ectodermal differentiation (18), was up-regulated in NP cells. The genes for the transcription factors *PAX6*, *POU3F2* (alias *BRN2*) and *ASCL1* (alias *MASH1*) were up-regulated in NP cells. *NR4A2* (alias *NURRI*), *MSX1* and *MYT1L*, which are involved in specification of DA neurons, were up-regulated during differentiation (Fig. 1B).

We examined whether the expression levels of *DNMTs* and *TETs* changed during differentiation using quantitative reverse-transcription PCR (qRT-PCR). Among the *DNMTs*, *DNMT3B* was highly expressed in hES cells, but its expression was reduced 119-fold in NP cells. Among *TETs*, *TET1* was highly expressed in hES cells, but its expression decreased 3.4-fold during differentiation; *TET2* and *TET3* expression increased during differentiation (Fig. 1C). The decreased expression of

*DNMT3B* and *TET1* led us to examine whether there was a global loss of DNA methylation and hydroxymethylation during differentiation. Using 5 mC- and 5 hmC-specific antibodies in enzyme-linked immunosorbent assays, we found that global levels of both 5 hmC and 5 mC significantly decreased during differentiation (Fig. 1D).

### Differentially hydroxymethylated or methylated regions associated with neural differentiation

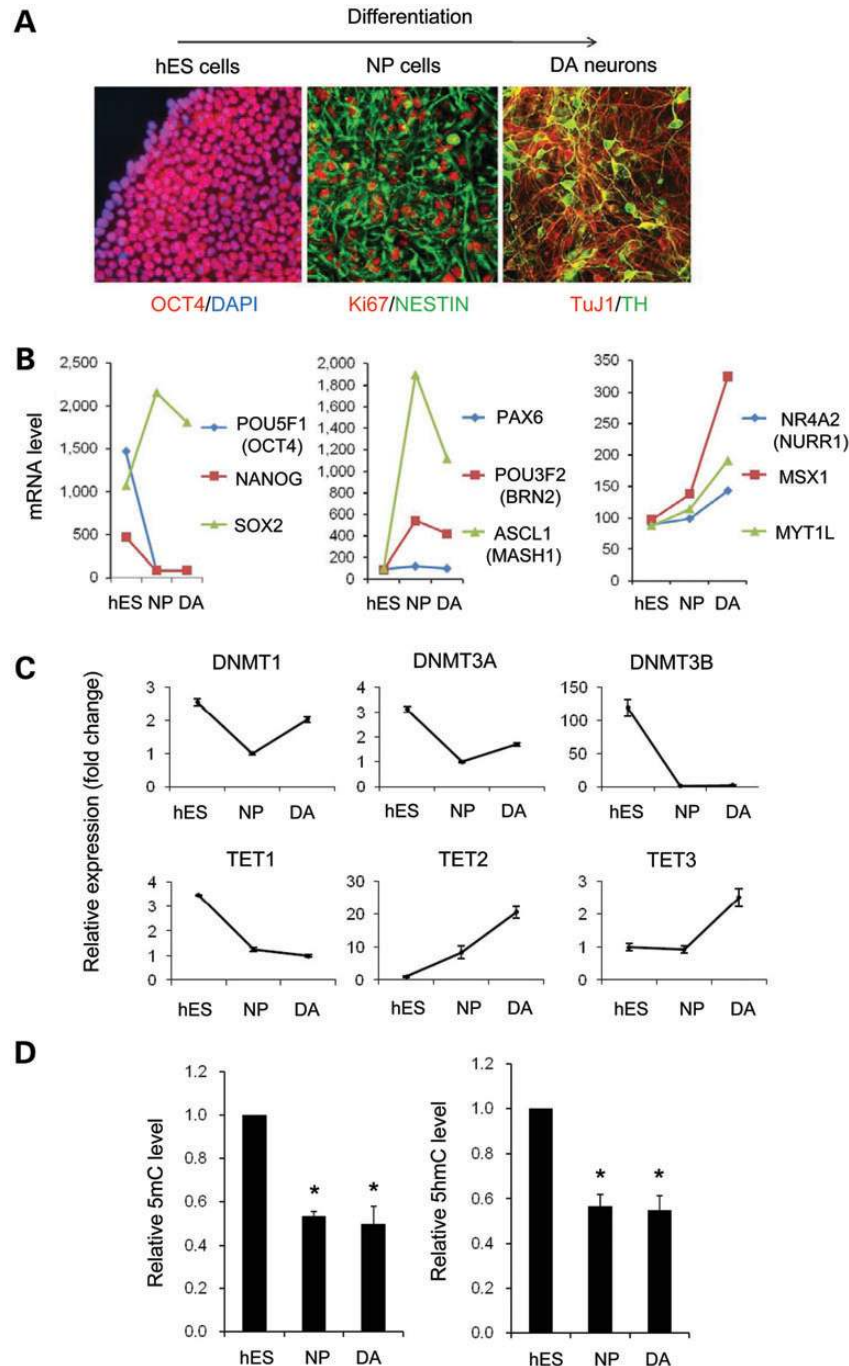
During differentiation of hES cells into NP cells, 946 genes, including 141 developmental process-related genes, were up-regulated > 2-fold, and 826 genes, including 187 DNA metabolism-related genes, were down-regulated >2-fold. During differentiation of NP cells into DA neurons, 219 genes, including 23 transport-related genes, were up-regulated and 236 genes, including 17 carbohydrate metabolism-related genes, were down-regulated (Fig. 2A).

MBD-seq is based on enrichment of methylated DNA in sequences. Thus MBD-seq is sensitive to regions that are highly methylated and have high CpG density (19). To evaluate the accuracy of our MBD-seq, we made a pair-wise comparison between methylation levels detected by whole-genome BS (downloaded from GEO with accession number GSM706061) and MBD-seq in hES cells at 1 893 389 CpG sites in 21 639 CpG islands (CGIs). We found significant correlation between the two methods ( $r = 0.62$ ,  $P < 2.2e-16$ ; Supplementary Material, Fig. S1).

To examine the global dynamics of DNA methylation and hydroxymethylation changes during differentiation, we first analyzed genome-wide changes in the distribution of 5 mC and 5 hmC. Using the model-based analysis of ChIP-Seq (MACS), a Poisson-based peak identification algorithm (20), we identified 99 238 5 mC-enriched regions and 115 913 5 hmC-enriched regions in MBD-seq and hMeDIP-seq data for hES cells, respectively. The change in the distribution of 5 mC was most pronounced during differentiation of hES cells into NP cells ( $r = 0.54$ ; Fig. 2B). Overall, the 5 hmC pattern was more stable than that of 5 mC; nevertheless, the 5 hmC pattern changed dramatically during differentiation of hES cells into NP cells ( $r = 0.68$ ) and during differentiation of NP cells into DA neurons ( $r = 0.78$ ; Fig. 2C). The duplicate biological experiments also revealed dramatic changes in the 5 mC and 5 hmC patterns during differentiation (Supplementary Material, Fig. S2).

Previous studies have shown that 5 hmC is enriched at enhancers (10,11) and exons (21). We used enhancer information (profiled by p300, H3K4me1 and H3K27ac in H9 hES cells) by downloading from GEO with accession number GSE24447 (22) and analyzed the genomic distribution of 5 hmC and 5 mC in hES cells. hMeDIP-seq showed that 5 hmC is enriched >4.6-fold above that expected by chance in enhancers. Exons also showed high enrichment of 5 hmC with >3.6-fold above that expected by chance (Fig. 2D left). MBD-seq revealed that 5 mC is enriched in exons (>7.0-fold) but less enriched in enhancers (>1.5-fold) than 5 hmC (Fig. 2D right).

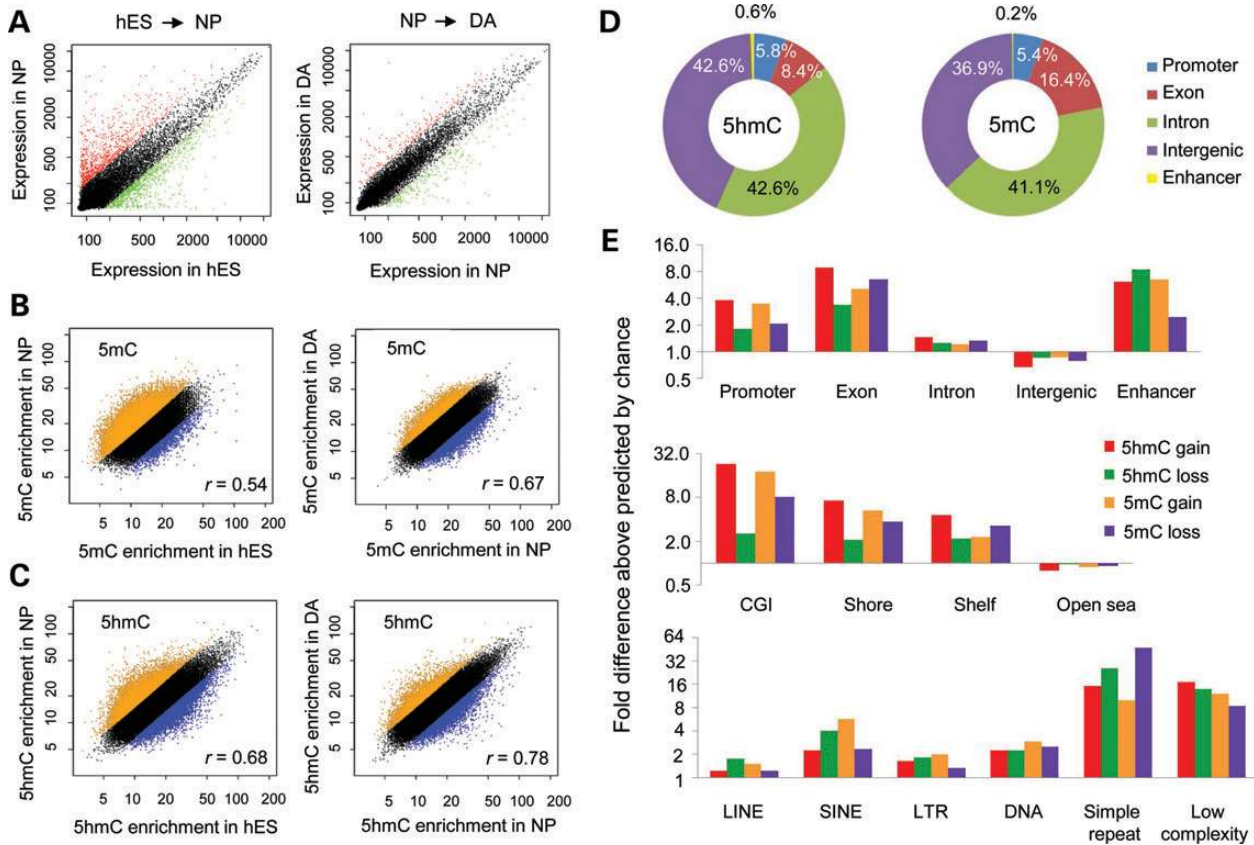
We identified 26 044 (22.6% of 115 913 5 hmC peaks) differentially hydroxymethylated regions (DhMRs,  $P < 10^{-5}$ ) and 16 123 (16.2% of 99 238 5 mC peaks) differentially methylated regions (DMRs,  $P < 10^{-5}$ ) during differentiation of hES cells into NP cells. Genomic distribution of DhMRs and DMRs



**Figure 1.** *In vitro* model system of differentiation of hES cells into NP cells and terminally into DA neurons. (A) Immunofluorescence staining of OCT4 (ES cell marker) in hES cells, NESTIN (NP cell marker)/Ki67 (proliferating cell marker) in NP cells and TH (DA neuron marker)/TuJ1 (neural marker) in DA neurons. (B) Relative expression of the hES cell markers *POU5F1*, *NANOG* and *SOX2*, the NP cell markers *PAX6*, *POU3F2* and *ASCL1*, and the DA neuron markers *NR4A2*, *MSX1* and *MYT1L* in the three differentiation states. (C) qRT-PCR of *DNMTs* and *TETs* in hES cells, NP cells and DA neurons ( $n = 3$ , values are the mean  $\pm$  standard error). (D) Relative global DNA methylation and hydroxymethylation levels in hES cells, NP cells and DA neurons measured by enzyme-linked immunosorbent assay with antibodies specific for 5 mC and 5 hmC ( $n = 6$ , values are the mean  $\pm$  standard error, \* $P < 0.05$ ).

during differentiation of hES cells to NP cells indicated that the most dramatic changes in 5 hmC/5 mC occurred in promoters (within 2 kb of the transcription start site (TSS)), exons and enhancers (Fig. 2E, upper panel). Notably, exons showed a gain of 5 hmC  $>8.9$ -fold above that expected by chance and a loss of 5 mC  $>6.6$ -fold above that expected by chance.

Among the DhMRs in enhancers, 131 showed a loss of 5 hmC and 106 showed a gain of 5 hmC. Pluripotency-related genes such as *NANOG* showed decreased 5 hmC in the closet enhancer, whereas *TH* showed increased 5 hmC in the closet enhancer during differentiation (Supplementary Material, Fig. S3). The duplicate biological experiments also revealed dynamic



**Figure 2.** Characterization of DhMRs within the hMeDIP-seq data and DMRs within the MBD-seq data. Scatter plots of gene expression (A), DNA methylation values (B) and DNA hydroxymethylation values (C) of the human genome in hES cells compared with NP cells and in NP cells compared with DA neurons. Red and yellow indicate a  $>2$ -fold increase, and green and blue indicate a  $>2$ -fold decrease. (D) Genomic distribution of 5 hmC- or 5 mC-enriched regions in hES cells. Genomic features were defined by the RefSeq database, and promoters were defined as  $\pm 2$  kb from the TSS annotated in the RefSeq database. Enhancers were identified by a chromatin signature (high p300, H3K4me1 and H3K27ac) in H9 hES cells (22) (E) Association of all DhMRs/DMRs identified with genomic features (upper), CGIs and neighboring context (middle), and repetitive sequences (lower). Shore,  $\sim 0$ – $2$  kb from the CGI; Shelf,  $\sim 2$ – $4$  kb from the CGI; Open Sea,  $>4$  kb from the CGI; SINE, short-interspersed nuclear element; LINE, long-interspersed nuclear element; LTR, long terminal repeat.

changes in 5 mC and 5 hmC patterns in promoters, exons and enhancers during differentiation (Supplementary Material, Fig. S4)

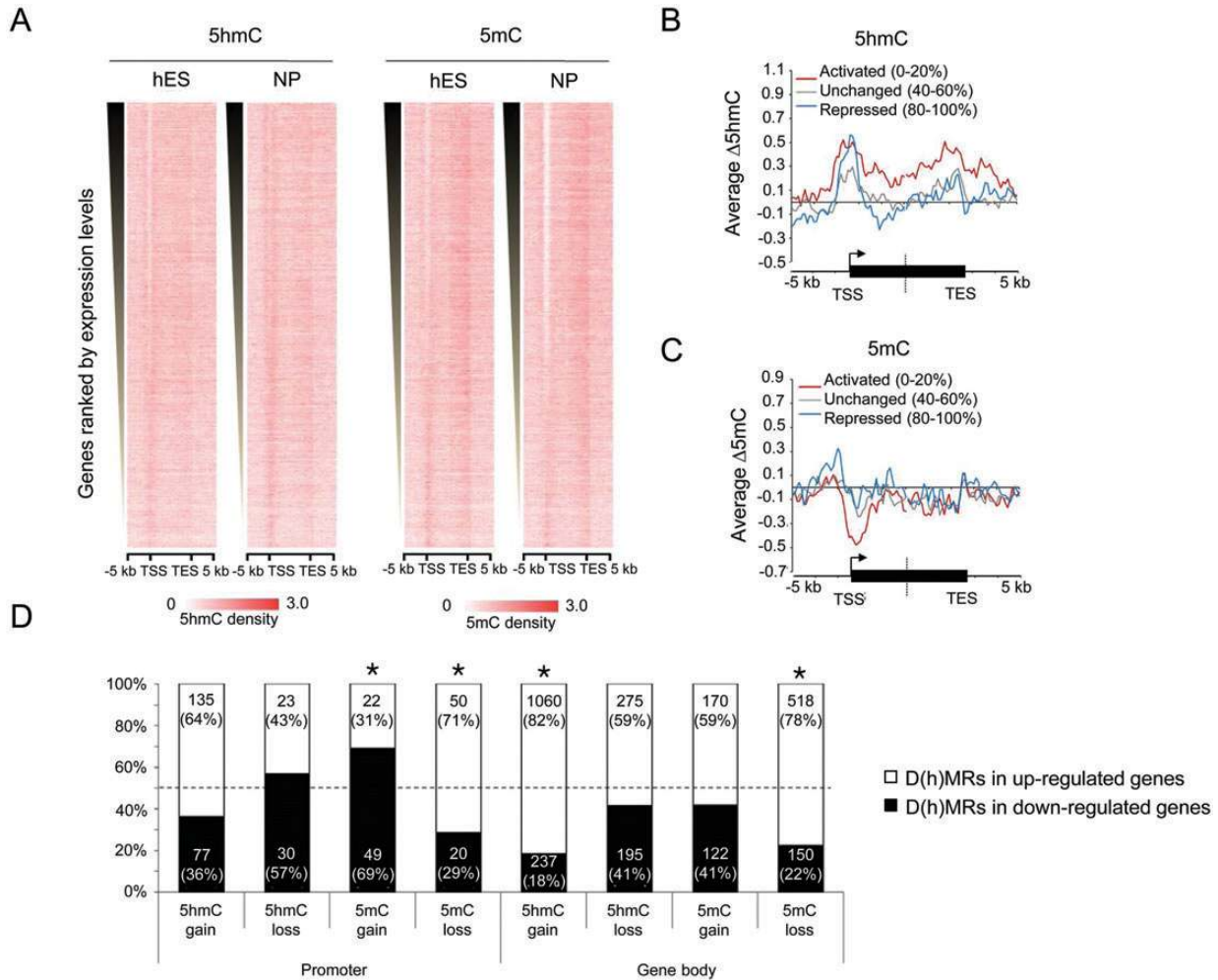
Although CGIs account for only about 0.7% of the human genome (23), we found dynamic changes in 5 hmC and 5 mC densities in CGIs and their shores ( $\sim 0$ – $2$  kb from the CGI) and shelf regions ( $\sim 2$ – $4$  kb from the CGI; Fig. 2E, middle panel). Gain of 5 hmC occurred in CGIs, shores and shelf regions  $>22.8$ -fold,  $7.2$ -fold and  $4.5$ -fold above that expected by chance, respectively. However, recent whole-genome single-base resolution 5 hmC maps of hES cells showed that 5 hmC is abundant in regions of low CpG content (11). Thus, it is possible that our affinity-based 5 mC/5 hmC sequencing methods, which may have bias toward high CpG density (19), may not sufficiently detect changes in the 5 mC/5 hmC distribution in regions of low CpG content.

Nearly half of the human genome is composed of repetitive sequences (24). We examined changes in 5 hmC and 5 mC densities at these repetitive sequence loci during differentiation. SINEs (short-interspersed nuclear elements) showed frequent loss of 5 hmC ( $>4.1$ -fold above that expected by chance) and a frequent gain of 5 mC ( $>5.8$ -fold above expected), suggesting that 5 hmC and 5 mC have different roles at SINEs during

differentiation of hES cells. Simple repeat and low-complexity sequences also showed dynamic changes in 5 hmC/5 mC during differentiation (Fig. 2E, lower panel). Taken together, these data suggested that DNA hydroxymethylation has both stable and dynamic features during differentiation.

#### Analysis of 5 hmC and 5 mC occurrence associated with gene expression

To investigate the potential role of DNA hydroxymethylation and methylation in controlling gene expression during differentiation, we ranked all of the genes identified in our transcriptional analysis by expression level in hES cells and in NP cells and then analyzed the distribution of 5 hmC or 5 mC within these genes. This analysis showed that the 5 hmC distribution was distinct from that of 5 mC (Fig. 3A, Supplementary Material, Fig. S5). During differentiation of hES cells into NP cells, activated genes acquired 5 hmC near the TSS and within the gene body, whereas repressed genes acquired 5 hmC only near the TSS (Fig. 3B). In the case of 5 mC, activated genes lost 5 mC in their TSS, whereas repressed genes acquired 5 mC upstream of the TSS during the differentiation of hES cells into NP cells (Fig. 3C). These data suggested that both 5 hmC and 5 mC



**Figure 3.** Gene expression-dependent 5 hmC and 5 mC densities in hES cells and NP cells. (A) Heatmap representation of 5 hmC (left) and 5 mC (right) densities within 5 kb of TSSs and TESs in hES cells and NP cells as a function of gene expression level. Average  $\Delta 5$  hmC (B) and  $\Delta 5$  mC (C) values within 5 kb of the TSS and TES for activated (top 20%), unchanged (middle 20%) or repressed (bottom 20%) genes. (D) Bar graph showing the association of D(h)MRs in the promoter ( $\pm 2$  kb of TSS) or gene body with up-regulated or down-regulated genes ( $*P < 0.05$ ).

play complex roles in gene expression involving both promoters and the gene body.

We also observed a statistically significant difference in 5 hmC/5 mC distribution between up-regulated and down-regulated genes, further supporting a role for 5 hmC/5 mC in regulating gene expression. Gain of 5 mC in promoters was significantly associated with down-regulated genes ( $P = 0.002$ ), whereas its loss in promoters was related to up-regulated genes ( $P = 0.010$ ). Notably, gain of 5 hmC within the gene body was significantly associated with up-regulated genes ( $P = 1.8 \times 10^{-5}$ ) and loss of 5 mC in the gene body was also associated with up-regulated genes ( $P = 0.003$ ; Fig. 3D).

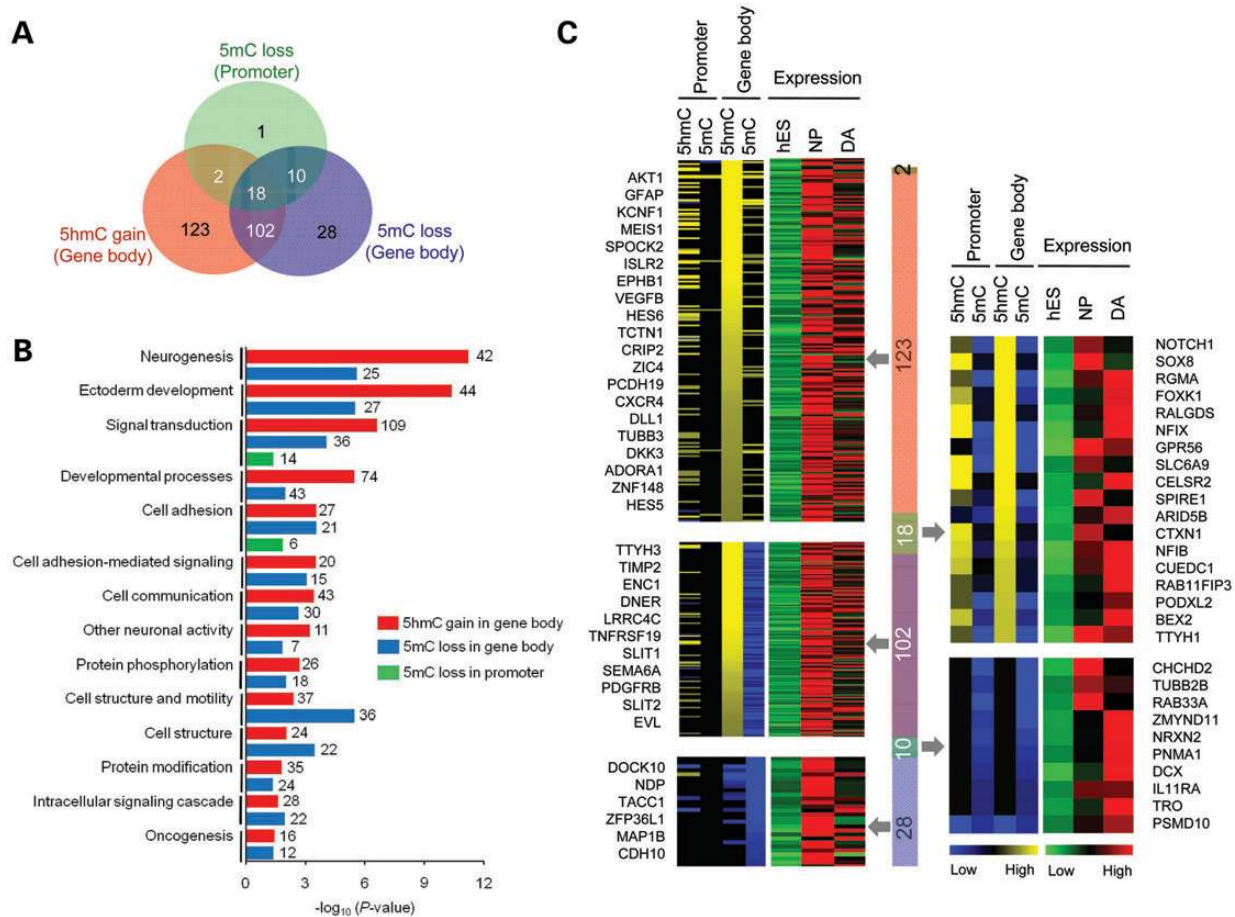
#### Gene body hydroxymethylation in genes up-regulated during neural differentiation

We next looked more closely at the hydroxymethylation of genes that were up-regulated during differentiation of hES cells into NP cells. Among the 946 genes that were up-regulated >2-fold, 31 had reduced 5 mC in the promoter, 245 gained

5 hmC in the gene body and 158 had reduced 5 mC in the gene body, with 30.4% overlap between the three groups (Fig. 4A, Supplementary Material, Table S1). Gene ontology analysis showed that the majority of these up-regulated genes are involved in neurogenesis, ectoderm development and signal transduction (Fig. 4B).

Increased expression during differentiation was observed for NOTCH1, a mediator of cell fate determination and regulator of NP cell maintenance and differentiation (25), RGMA, an axon guidance molecule involved in central nervous system development (26), and FOXP1, a forkhead family transcription factor important in embryonic development (27) (Figs 4C and 5A). These three genes gained 5 hmC in the gene body and lost 5 mC in the promoter and gene body (Figs 4C and 5B).

Similarly, the genes for AKT1, a serine-threonine kinase that is important for NP cell proliferation and inhibition of differentiation (20), and ISLR2 (alias LINX), a leucine-rich repeat and immunoglobulin family protein that is expressed in developing sensory and motor neurons (28), showed 5 hmC enrichment in the gene body during differentiation (Figs 4C and 5B).



**Figure 4.** Up-regulated genes that gained 5 hmC in the gene body or lost 5 mC in the promoter or gene body. (A) Venn diagram showing overlap of up-regulated genes (>2-fold) that lost 5 mC in the promoter with those that gained 5 hmC in the gene body, and those that lost 5 mC in the gene body. (B) Gene ontology analysis for the genes of the three categories. (C) Heatmap representation of differences in 5 mC and 5 hmC densities of D(h)MRs in promoters and gene bodies (left). Heatmap of relative gene expression levels in hES cells, NP cells and DA neurons (right) grouped by 5 mC/5 hmC gain or loss as in the Venn diagram.

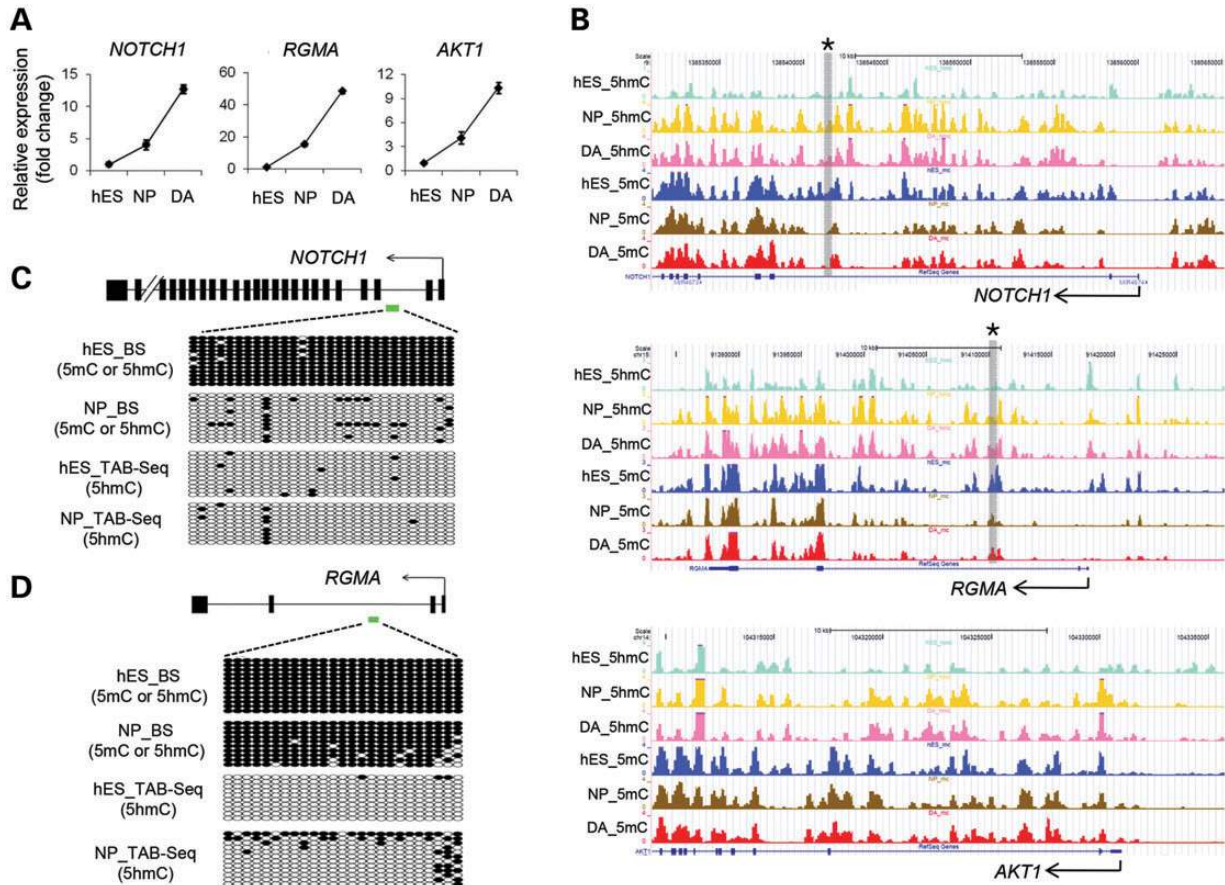
In addition, the NOTCH1 signaling pathway genes *DLL1*, *HES5*, *DNER* and *GFAP* also gained 5 hmC in the gene body (Fig. 4C).

To refine the differentiation-induced changes in 5 mC and 5 hmC that we observed to single-base resolution, we performed locus-specific BS and TAB-Seq. We first tested the efficiency of both methods using 100% 5 hmC and 100% 5 mC control DNA. In our hands, BS detected 99.8% of 5 mC and 99.5% of 5 hmC, whereas TAB-Seq detected 1.5% of 5 mC and 96.9% of 5 hmC (Supplementary Material, Fig. S6). We next performed BS and TAB-Seq on 29 CpG sites within the CGI located in the *NOTCH1* second intron. Combining results from the two sequencing methods, we found that 28 CpG sites were methylated in hES cells and demethylated in NP cells, whereas one CpG site was methylated in hES cells and was hydroxymethylated in NP cells (Fig. 5C). In the case of 25 CpG sites within the CGI located in the *RGMA* second intron, most CpG sites were methylated in hES cells, whereas demethylation and hydroxymethylation slightly increased in NP cells (Fig. 5D). These data indicated that 5 hmC not only is an intermediate of DNA demethylation but also exhibits substantial stability—enough so that its occurrence can be readily detected in the genome.

### DNA methylation as a locking mechanism to maintain gene silencing

Promoter methylation is a major epigenetic mechanism for initiating gene silencing and also acts as a locking mechanism to prevent reactivation of downstream genes (29). Among the 554 genes that gained 5 mC in their promoter regions during differentiation, 35 were down-regulated >2-fold (Fig. 6A). The genes for the stem cell markers *POU5F1* and *ZFP42* (alias *REX1*) (30) acquired 5 mC in their promoters. Similarly, the genes for *HMGA1* and *HMGB3*, which are chromatin remodeling proteins involved in stem cell transcriptional networks (27), and *PARP1*, a multifunctional regulator of chromatin structure and transcription (31), were hypermethylated and down-regulated (Fig. 6B).

We found that 358 genes acquired DNA methylation during neural differentiation although they were already silent in hES cells (signal intensity < 200; fold change in expression between 0.8 and 1.2; Fig. 6A and Supplementary Material, Table S2). We refer to these genes as 'locked genes'. Gene ontology analysis indicated that locked genes were often associated with cell communication, signal transduction and ion transport. Members of



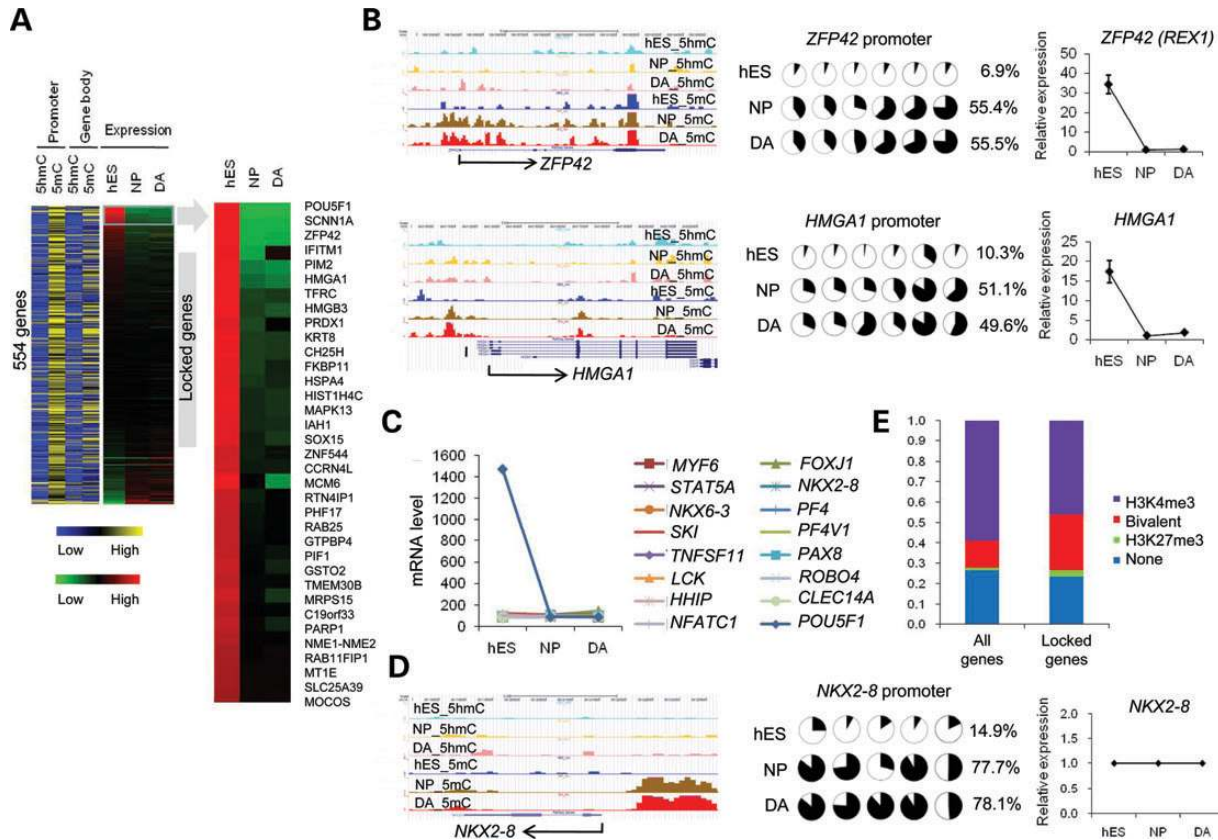
**Figure 5.** Distribution of 5 hmC and 5 mC densities in *NOTCH1*, *RGMA* and *AKT1*. (A) qRT-PCR of *NOTCH1*, *RGMA* and *AKT1* in hES cells, NP cells and DA neurons ( $n = 3$ , values are the mean  $\pm$  standard error). (B) hMeDIP-seq and MBD-seq results for *NOTCH1*, *RGMA* and *AKT1*. \*Gray shading indicates target regions for TAB-Seq. (C) BS and TAB-Seq results for 29 CpG sites within the CGI located in the *NOTCH1* second intron. (D) BS and TAB-Seq results for 25 CpG sites within the CGI located in the *RGMA* second intron. Each circle represents the cytosine of a CpG site. Open circles, sequenced as T; filled circles, sequenced as C. Each row represents a single clone.

several gene family clusters, such as zinc finger proteins (24 genes), were included in the locked genes. Interestingly, the locked genes included 15 mesoderm development-related genes (Fig. 6C). For example, *NKX2-8*, a developmentally regulated transcription factor (32), acquired promoter methylation and remained silenced throughout the neural differentiation of hES cells (Fig. 6D). These data suggested that promoter methylation of mesoderm development-related genes maintains gene silencing during ectodermal differentiation of ES cells. Genes containing the promoter type ‘bivalent’ or ‘histone 3 with trimethylated lysine 27’ were enriched in the locked gene group (Fig. 6E), suggesting that locked genes may be repressed in hES cells by a separate mechanism, such as binding of the polycomb-group proteins, and then acquire DNA methylation during neural differentiation.

## DISCUSSION

5hmC is considered an intermediate in DNA demethylation (5,7), but its role in regulating transcription has been controversial. Hydroxymethylation is enriched in both the gene body of highly expressed genes and the TSS of poorly expressed genes (33). Like 5 mC, 5 hmC in promoters is associated with reduced gene expression (34) but can also partially overcome

the silencing effect of 5 mC (35). However, these previous findings were obtained using pluripotent mouse ES cells. To examine the role of 5 hmC in gene expression during differentiation, we compared 5 hmC patterns of NP cells with that of undifferentiated hES cells. During neural differentiation of hES cells, both activated genes and repressed genes had increased 5 hmC signals in the TSS, but only activated genes showed increased 5 hmC in the gene body (Fig. 3B). This is consistent with a previous report of increased 5 hmC at promoters and in the gene body of activated genes during postnatal neurodevelopment of mouse cerebellum (36). It was also recently reported that binding of methyl-CpG-binding protein 2 (MeCP2) to 5 hmC facilitates transcription in neural cell types (37). It has been suggested that Tet1, which converts 5 mC to 5 hmC, has dual functions in transcriptional regulation (38), i.e. mediating promoter hypomethylation to maintain expression of activated genes and associating with the Sin3A co-repressor complex to repress a subset of target genes (39). We suggest that, like Tet1, 5 hmC within gene promoters may have a dual function in transcriptional regulation, whereas 5 hmC within the gene body appears to be mainly involved in gene activation. The detailed mechanisms of these parallel functions have yet to be elucidated. TAB-Seq of several gene body regions showed



**Figure 6.** Characterization of the promoter-methylated genes during neural differentiation of hES cells. (A) Heatmap representations of differences in 5 hmC and 5 mC densities of D(h)MRs in promoters and gene bodies (left) and relative gene expression levels in hES cells, NP cells and DA neurons (right). (B) hMeDIP-seq and MBD-seq, Infinium 450 BeadChip analysis, and qRT-PCR analysis of *ZFP42* (upper) and *HMGA1* (lower). Each pie chart shows the average methylation of a particular CpG site from duplicate experiments. (C) Expression of 15 mesoderm development-related genes and *POU5F1* during differentiation. (D) hMeDIP-seq and MBD-seq, Infinium 450 K analysis and qRT-PCR analysis of *NKX2-8*. (E) Correlation between locked genes and promoter histone modification groups. The four promoter classification groups are based on a published set of histone modifications in H9 hES cells: H3K4me3-only, Bivalent (H3K4me3 and H3K27me3), H3K27-only and None (45).

that 5 hmC was retained at certain sites having a specific sequence during demethylation (Fig. 5C) or was deposited over relatively broad regions (Fig. 5D). A recent report revealed factors that serve as dynamic readers of 5 hmC during differentiation (40). 5 hmC recruits distinct transcription regulators as well as DNA repair proteins for active demethylation (40). We suggest that 5 hmC in the gene body may be an important signal for active transcription by recruiting specific transcription regulators or recruiting MeCP2 during neural differentiation.

Although we found that promoter methylation was not always sufficient to repress transcription, we also found that several pluripotency-related genes acquired promoter methylation and were repressed during differentiation. Interestingly, the chromatin architecture-related genes *HMGA1*, *HMGB3* and *PARP1* were repressed by promoter methylation during differentiation. *HMGA1* induces the expression of the pluripotency genes *SOX2*, *LIN28* and *cMYC* and enhances cellular reprogramming of somatic cells to pluripotent stem cells (27). *Parp1* also appears to play a key role in early-stage epigenetic modification during somatic cell reprogramming (41). We suggest that repression of *HMGA1* and *PARP1* by promoter methylation may alter the transcriptional network of pluripotent hES cells to promote neural differentiation.

This study identified 358 genes that were locked during neural differentiation of hES cells. These genes were silent in hES cells despite having a relatively low density of 5 mC in their promoters. We hypothesize that other repressive mechanisms, such as binding of the polycomb-group proteins, may be involved in the silencing of these genes in hES cells. Although silent in hES cells, these genes acquired 5 mC in their promoters during differentiation. Many of these locked genes are involved in cell communication and differentiation. Notably, genes related to mesoderm development acquired promoter methylation during neural differentiation of hES cells. These data support previous observations that DNA methylation is an epigenetic mechanism not only for initiating gene silencing but also for preventing gene reactivation (29).

## MATERIALS AND METHODS

### Differentiation of hES cells

Differentiation of the hES cell line H9 (WiCell; passages 35–38) into NP cells and terminally into DA neurons was performed as described (17).



### qRT-PCR

qRT-PCR was performed as described (42). PCR primer sequences are listed in Supplementary Material, Table S3.

### 5 hmC and 5 mC quantification

The 5 hmC and 5 mC content in cells was quantified using the MethylFlash hydroxymethylated DNA quantification kit and the MethylFlash methylated DNA quantification kit (Epigenetek). Both kits are based on enzyme-linked immunosorbent assays.

### Expression analysis

Gene expression in hES cells, NP cells and DA neurons was assayed in duplicate. RNA isolated from cells at various time points was used for gene expression analysis using the Human-6 Whole-Genome Expression BeadChip (Illumina) as described (15).

### hMeDIP-seq and MBD-seq

Genomic DNA was isolated with a DNeasy Blood and Tissue kit (QIAGEN) and fragmented to 150–300 bp by 60 PSI nitrogen gas for 7 min in a nebulizer (Illumina), and then a library was prepared using a genomic DNA sample prep kit (Illumina). Briefly, 5  $\mu$ g of fragmented DNA was end-repaired, A-tailed and ligated to paired-end adapters. The pre-adapted 5 hmC containing DNA was immunoprecipitated using an hMeDIP kit (Diagenode). Briefly, 1  $\mu$ g of the pre-adapted DNA was denatured for 10 min at 95°C and incubated with 2.5  $\mu$ g of mouse monoclonal anti-5 hmC (Diagenode) for 2 h at 4°C. The immunoabsorbed DNA was recovered with anti-mouse IgG-coated magnetic beads (Diagenode) at 4°C overnight. The immunoprecipitated DNA was amplified by 18 cycles of PCR using Illumina PCR primers, and the products were size-fractionated on a 2% agarose gel to obtain fragments of 200–300 bp. Cluster generation and 76 cycles of single-read sequencing were then performed using Illumina Genome Analyzer Iix. Input genomic DNA and mouse IgG controls were also sequenced. MBD-seq was performed as described (15).

### BS and TAB-Seq

For BS, genomic DNA (1  $\mu$ g) was modified with sodium bisulfite using the EZ DNA Methylation-Gold kit (Zymo Research). For TAB-Seq, genomic DNA (1  $\mu$ g) was glucosylated by  $\beta$ -glucosyltransferase and oxidized by Tet1 using the 5 hmC TAB-Seq kit (WiseGene) and then modified by sodium bisulfite using EZ DNA Methylation-Gold kit. We used the 5 mC and 5 hmC DNA Set (Zymo Research) as standards. Genes of interest were amplified by PCR using the primers listed in Supplementary Material, Table S3. The PCR products were purified and cloned using the pGEM-T Easy Vector system (Promega). Twelve clones for each gene were randomly chosen for sequencing.

### Gene expression-dependent 5 hmC and 5 mC density analysis

Heatmaps of gene expression-dependent 5 hmC and 5 mC densities were constructed from the data acquired from the gene expression array, hMeDIP-Seq and MBD-Seq. We sorted genes by expression levels in hES and NP cells and then estimated the 5 hmC and 5 mC densities in 200 bp-sized bins within 5 kb of the TSS and within 5 kb of the transcription end site (TES) of each gene using the refGene annotation of UCSC Genome Bioinformatics (<http://genome.ucsc.edu/>). A normalized methylation enrichment score for 5 hmC or 5 mC was calculated from the number of mapped 5 hmC and 5 mC reads in each bin divided by the bin size (200 bp) and the total number of mapped sequence reads on the genome for each hES or NP cell multiplied by the human genome size ( $3 \times 10^8$  bp).

Variations in 5 hmC and 5 mC in three gene expression groups were also plotted based on the gene expression data and sequencing read densities in hES cells and NP cells. First, genes were sorted by the difference in expression between hES and NP cells and then divided into activated (top 20%), unchanged (middle 20%) and repressed (bottom 20%) groups. Meta-gene profiles of 5 hmC and 5 mC enrichment within 5 kb of the TSS and TES were calculated for the three gene expression groups by averaging the group-wide 5 hmC or 5 mC read densities of each 200 bp-sized bins in these regions.

### DMR and DhMR identification and annotation

The effects of methylation or hydroxymethylation in the gene body and promoter on gene expression were analyzed by comparing the promoter-associated and gene body-associated DMRs or DhMRs with gene expression values. We mapped MBD-Seq and hMeDIP-Seq reads on the human reference genome (hg18 of the UCSC genome) using the Bowtie software (43) allowing only unique aligned reads with less than three mismatches. We then identified DMRs and DhMRs using the MACS software (20) with the default options, including  $P$ -value  $< 10^{-5}$ . NP cell-originated sequence data were considered as cases and hES cell-originated sequence data were considered as controls in the analysis. We selected significant positive peaks and significant negative peaks from the MACS output and determined promoter-associated or gene body-associated DMRs or DhMRs based on the positional overlaps between DMRs or DhMRs and the promoter regions (TSS  $\pm$  2 kb) or gene body regions of the reference gene mRNAs.

### Methylation analysis using the infinium human methylation 450 BeadChip

Methylation of hES cells, NP cells and DA neurons was assayed in duplicate. Genomic DNA (1  $\mu$ g) from each sample was bisulfite converted using the EZ DNA Methylation-Gold kit (Zymo Research), and 200 ng of the converted DNA was used for PCR amplification. Amplified DNA was hybridized on the Infinium HumanMethylation 450 BeadChip (Illumina) following the Illumina Infinium HD Methylation protocol. The arrays were imaged using an Illumina HiScan SQ scanner, and images were processed and intensity data extracted according to the manufacturer's instructions. Each methylation signal

was used to compute the  $\beta$  value, which is a quantitative measure of DNA methylation ranging from 0 (completely unmethylated cytosines) to 1 (completely methylated cytosines) (44).

### Data access

Microarray and sequencing data have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE38217.

### SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

*Conflict of Interest statement.* None declared.

### FUNDING

This work was supported by grants from the Stem Cell Research Program (2012M3A9B4027954), the Future-Based Technology Development Program (NRF2011-0015710) and the KRIBB Research Initiative Program, which are funded by the Korean Ministry of Education, Science and Technology.

### REFERENCES

- Reik, W. (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, **447**, 425–432.
- Okano, M., Xie, S. and Li, E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.*, **19**, 219–220.
- Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R. and Fan, G. (2008) Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell*, **2**, 160–169.
- Walsh, C.P., Chaillet, J.R. and Bestor, T.H. (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.*, **20**, 116–117.
- Wu, S.C. and Zhang, Y. (2010) Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.*, **11**, 607–620.
- Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
- Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C. and Zhang, Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1133.
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L. and Song, H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, **145**, 423–434.
- Serandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Paliere, G., Gheeraert, C., Barloy-Hubler, F., Peron, C.L. et al. (2012) Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res.*, **40**, 8255–8265.
- Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B. et al. (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*, **149**, 1368–1380.
- Meissner, A. (2010) Epigenetic modifications in pluripotent and differentiated cells. *Nat. Biotechnol.*, **28**, 1079–1088.
- Park, C.H., Minn, Y.K., Lee, J.Y., Choi, D.H., Chang, M.Y., Shim, J.W., Ko, J.Y., Koh, H.C., Kang, M.J., Kang, J.S. et al. (2005) In vitro and in vivo analyses of human embryonic stem cell-derived dopamine neurons. *J. Neurochem.*, **92**, 1265–1276.
- Perrier, A.L., Tabar, V., Barberi, T., Rubio, M.E., Bruses, J., Topf, N., Harrison, N.L. and Studer, L. (2004) Derivation of midbrain dopamine neurons from human embryonic stem cells. *Proc. Natl Acad. Sci. USA*, **101**, 12543–12548.
- Kim, M., Kang, T.W., Lee, H.C., Han, Y.M., Kim, H., Shin, H.D., Cheong, H.S., Lee, D., Kim, S.Y. and Kim, Y.S. (2011) Identification of DNA methylation markers for lineage commitment of *in vitro* hepatogenesis. *Hum. Mol. Genet.*, **20**, 2722–2733.
- Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. and Jacobsen, S.E. (2011) 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.*, **12**, R54.
- Ko, J.Y., Park, C.H., Koh, H.C., Cho, Y.H., Kyhm, J.H., Kim, Y.S., Lee, I., Lee, Y.S. and Lee, S.H. (2007) Human embryonic stem cell-derived neural precursors as a continuous, stable, and on-demand source for human dopamine neurons. *J. Neurochem.*, **103**, 1417–1429.
- Thomson, M., Liu, S.J., Zou, L.N., Smith, Z., Meissner, A. and Ramanathan, S. (2011) Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell*, **145**, 875–889.
- Bock, C., Tomazou, E.M., Brinkman, A.B., Muller, F., Simmer, F., Gu, H., Jager, N., Gnirke, A., Stunnenberg, H.G. and Meissner, A. (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.*, **28**, 1106–1114.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
- Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S. et al. (2011) Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell*, **42**, 451–464.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A. and Wysocka, J. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, **470**, 279–283.
- Jelinek, J., Liang, S., Lu, Y., He, R., Ramagli, L.S., Shpall, E.J., Estecio, M.R. and Issa, J.P. (2012) Conserved DNA methylation patterns in healthy blood cells and extensive changes in leukemia measured by a new quantitative technique. *Epigenetics*, **7**, 1368–1378.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
- Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.G., Jiang, Y., Pfeifer, G.P. et al. (2013) Dynamics of 5-hydroxymethylcytosine and chromatin marks in mammalian neurogenesis. *Cell Rep.*, **3**, 291–300.
- Matsunaga, E., Tauszig-Delamasure, S., Monnier, P.P., Mueller, B.K., Strittmatter, S.M., Mehlen, P. and Chedotal, A. (2004) RGM and its receptor neogenin regulate neuronal survival. *Nat. Cell Biol.*, **6**, 749–755.
- Carlsson, P. and Mahlapuu, M. (2002) Forkhead transcription factors: key players in development and metabolism. *Dev. Biol.*, **250**, 1–23.
- Mandai, K., Guo, T., St Hillaire, C., Meabon, J.S., Kanning, K.C., Bothwell, M. and Ginty, D.D. (2009) LIG Family receptor tyrosine kinase-associated proteins modulate growth factor signals during neural development. *Neuron*, **63**, 614–627.
- Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T. and Weber, M. (2010) Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.*, **42**, 1093–1100.
- Patterson, M., Chan, D.N., Ha, I., Case, D., Cui, Y., Handel, B.V., Mikkola, H.K. and Lowry, W.E. (2011) Defining the nature of human pluripotent stem cell progeny. *Cell Res.*, **22**, 178–193.
- Krishnakumar, R. and Kraus, W.L. (2010) The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. *Mol. Cell*, **39**, 8–24.
- Kajiyama, Y., Tian, J. and Locker, J. (2002) Regulation of alpha-fetoprotein expression by Nkx2.8. *Mol. Cell Biol.*, **22**, 6122–6130.
- Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E. and Zhang, Y. (2011) Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.*, **25**, 679–684.
- Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P. et al. (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, **473**, 394–397.

35. Ficiz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S. and Reik, W. (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, **473**, 398–402.
36. Szulwach, K.E., Li, X., Li, Y., Song, C.X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I. *et al.* (2011) 5-hmC-mediated Epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.*, **14**, 1607–1616.
37. Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S. and Heintz, N. (2012) MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*, **151**, 1417–1430.
38. Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E. and Zhang, Y. (2011) Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*, **473**, 389–393.
39. Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappilber, J. and Helin, K. (2011) TET1 And hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, **473**, 343–348.
40. Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F. *et al.* (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, **152**, 1146–1159.
41. Doege, C.A., Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A. *et al.* (2012) Early-stage epigenetic modification during somatic cell reprogramming by Pcp1 and Tet2. *Nature*, **488**, 652–655.
42. Kim, M., Kim, J.H., Jang, H.R., Kim, H.M., Lee, C.W., Noh, S.M., Song, K.S., Cho, J.S., Jeong, H.Y., Hahn, Y. *et al.* (2008) LRRC3B, Encoding a leucine-rich repeat-containing protein, is a putative tumor suppressor gene in gastric cancer. *Cancer Res.*, **68**, 7147–7155.
43. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
44. Bibikova, M., Chudin, E., Wu, B., Zhou, L., Garcia, E.W., Liu, Y., Shin, S., Plaia, T.W., Auerbach, J.M., Arking, D.E. *et al.* (2006) Human embryonic stem cells have a unique epigenetic signature. *Genome Res.*, **16**, 1075–1083.
45. Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S. *et al.* (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.*, **4**, e1000242.