Cytosine modifications modulate the chromatin architecture of transcriptional enhancers

Elise A. Mahé, Thierry Madigou, Aurélien A. Sérandour, Maud Bizot, Stéphane Avner, Frédéric Chalmel, Gaëlle Palerme, Raphaël Métivier, and Gilles Salbert

1CNRS UMR6290, Equipe SP@RT, Institut de Génétique et Développement de Rennes, Campus de Beaulieu, 35042 Rennes Cedex, France; 2Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France; 3EMBL, Heidelberg 69117, Germany; 4Inserm U1085-IRSET, Université de Rennes 1, F-35042 Rennes, France

Epigenetic mechanisms are believed to play key roles in the establishment of cell-specific transcription programs. Accordingly, the modified bases 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) have been observed in DNA of genomic regulatory regions such as enhancers, and oxidation of 5mC into 5hmC by Ten-eleven translocation (TET) proteins correlates with enhancer activation. However, the functional relationship between cytosine modifications and the chromatin architecture of enhancers remains elusive. To gain insights into their function, 5mC and 5hmC levels were perturbed by inhibiting DNA methyltransferases and TETs during differentiation of mouse embryonal carcinoma cells into neural progenitors, and chromatin characteristics of enhancers bound by the pioneer transcription factors FOXA1, MEIS1, and PBX1 were interrogated. In a large fraction of the tested enhancers, inhibition of DNA methylation was associated with a significant increase in monomethylation of H3K4, a characteristic mark of enhancer priming. In addition, at some specific enhancers, 5mC oxidation by TETs facilitated chromatin opening, a process that may stabilize MEIS1 binding to these genomic regions.

[Supplemental material is available for this article.]

Cell differentiation relies on the engagement of short genomic regions called transcriptional enhancers, acting at a distance from the transcription start site of their target genes. Enhancers are endowed with the capacity to bind numerous cell-specific and general transcription factors (TFs) in a temporal sequence which is partly dictated by the timely controlled expression of TFs and/or their regulated binding to chromatin through signaling pathways (Spitz and Furlong 2012; Heinz et al. 2015). In their inactive state, enhancers can be targeted by so-called “pioneer factors” (PFs) which bind condensed chromatin and favor transition of enhancers to a primed state, in turn facilitating binding of additional TFs and further activation (Iwafuchi-Doi and Zaret 2014). Several TFs, including FOXA1, SPI1, PBX1, CEBPB, the AP-1 complex, and MEIS1, have been qualified as PFs and play major roles in pluripotent cell differentiation during development and in adulthood (Heidt et al. 2007; Lupien et al. 2008; Ghisletti et al. 2010; Biddie et al. 2011; Magnani et al. 2011; Siersbæk et al. 2011). PF engagement during lineage commitment is thought to provoke changes in enhancer chromatin architecture. In agreement with this view, binding of SPI1 to enhancer chromatin during monocytic-to-osteoclast differentiation impacts cytosine methylation (5mC) at CpGs through the recruitment either of DNA methyltransferases (DNMTs) at active enhancers, or of Ten-eleven translocation (TET) enzymes, which catalyze 5mC oxidation into 5-hydroxymethylcytosine (5hmC) (Tahiliani et al. 2009), at inactive enhancers (de la Rica et al. 2013). 5hmC is a catalytic intermediate in a pathway of active DNA demethylation generating further oxidative derivatives (5-formylcytosine [5fC] and 5-carboxylcytosine [5caC]) subsequently replaced by cytosines through action of the base excision repair (BER) machinery (He et al. 2011; Maiti and Drohat 2011). However, due to the relative stability of 5hmC and its ability to bind specific proteins (Iurlaro et al. 2013; Spruijt et al. 2013; Bachman et al. 2014), 5hmCpGs may also act as signaling modules just as SmCpGs do (Bird 2011). In this respect, high levels of 5hmC were found at selected enhancers prior to binding of the 3-aa loop extension-homeodomain (TALE-HD) proteins MEIS1 and PBX1 during differentiation of mouse P19 embryonal carcinoma cells (ECCs) into neural progenitors (Sérandour et al. 2012), suggesting that DNA hydroxymethylation at enhancers could favor MEIS1 and PBX1 engagement. Moreover, consistent with data showing that TF binding to regulatory regions lowers their CpG methylation levels (stadler et al. 2011), we previously observed that FOXA1 recruitment to enhancers is associated with DNA demethylation during P19 ECC differentiation (Sérandour et al. 2011). Nonetheless, it remains unclear which steps of enhancer priming and activation are regulated through DNA demethylation, and in particular, if 5hmC plays a role per se in these processes. In order to gain insights into the role of cytosine modifications at PF-bound enhancers, we have here analyzed the impact of DNA methylation/demethylation processes on the chromatin structure of FOXA1- and TALE-HD-bound enhancers.
Results

P19 ECCs as a model-system to study enhancer activation by pioneer factors during NPC specification

Although different from the embryonic stem cell (ESC) model classically used to analyze neural progenitor cell (NPC) commitment over a 6-d process, P19 ECCs, after treatment with retinoic acid (RA), have the ability to generate NPCs showing a hydroxymethylome similar to the one observed in ESC-derived NPCs, in a rapid (2 d) and robust fashion (Supplemental Fig. S1A–C). Importantly, 82.5% of the regions gaining 5hmC during P19 cell differentiation overlap with hydroxymethylated regions from ESC-derived NPCs. Furthermore, 5hmC-positive regions from ECC-derived NPCs identify 28 out of the 48 enhancers, as demonstrated by Visel et al. (2007) to be active in neural tissues from E11.5 dpc mouse embryos (Supplemental Fig. S1D,E; http://enhancer.lbl.gov/). Similarly, 5hmC-positive regions from ESC-derived NPCs identify 22 out of these 48 enhancers. Moreover, it has been demonstrated that P19 ECC differentiation into NPCs relies on the induction of FOXA1, PBX1, and MEIS1, three transcription factors which are also induced by RA in mouse ESCs (Kashyap et al. 2013), and is enhanced by their overexpression (Qin et al. 2004; Tan et al. 2010; Dong et al. 2012; Yamada et al. 2013). These observations indicate (1) that the hydroxymethylome of ECC-derived NPCs is not aberrant, and (2) that ECC-derived NPCs constitute a highly relevant cellular model to study the activation of enhancers related to developmental programs controlled by pioneer factors.

In this context, we previously reported a major reconfiguration of the hydroxymethylome of P19 ECCs upon RA treatment (Sérandour et al. 2012), which we show here to be correlated with loss of pluripotency factors NANOG and POU5F1 and activation of the pioneer factor genes FoxA1 and Meis1 (Fig. 1A,B). MEIS1 is known to associate with PBX1 to form a prevalent TALE-HD complex (Chang et al. 1997) which is likely to control P19 differentiation (Qin et al. 2004; Yamada et al. 2013). Accordingly, data obtained by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) indicate that MEIS1 and PBX1 may directly control the expression of NPC marker genes like Tshz1, Pax6, and Irx3 (Fig. 1C). In addition to the role of MEIS1/ PBX1 dimers, FOXA1 might also control P19 cell fate decision. Indeed, FOXA1 is rapidly induced in response to RA (Fig. 1B) and is known to stimulate NPC commitment in P19 cells (Tan et al. 2010; Dong et al. 2012). To investigate the relationship between FOXA1 and TALE-HDs, we mapped FOXA1 binding sites in P19-derived NPCs. A significant fraction of FOXA1 sites were also enriched in MEIS1 and PBX1 (Fig. 1D–G), suggesting an interplay between these pioneer factors in mediating chromatin activation at specific sites such as within the HoxB cluster during NPC lineage commitment.

Pioneer factor-bound enhancers are enriched in 5hmC

Primed enhancers are characterized by the presence of histone H3 monomethylated on lysine 4 (H3K4me1) and the absence of acetylated H3K27 (H3K27ac), whereas active enhancers present both marks (for review, see Calo and Wysocka 2013). We previously observed that, during RA-induced differentiation of P19 ECCs into NPCs, differentially hydroxymethylated regions (DhMRs) are detected at enhancers either being silenced for down-DhMRs or undergoing activation for up-DhMRs (Sérandour et al. 2012). In order to obtain a more complete view of these dynamic processes at the full sets of up- and down-DhMRs, we first analyzed the temporal variations of H3K4me1 and H3K27ac histone modifications genome-wide by ChIP-seq, together with those of 5hmC and 5mC by hydroxymethylated or methylated DNA immunoprecipitation (hMeDIP-seq and MeDIP-seq, respectively) during RA-induced differentiation of P19 ECCs. In addition, chromatin opening was assessed by formaldehyde assisted isolation of regulatory elements (FAIRE) (Giresi et al. 2007). For both types of DhMRs, the levels of H3K4me1 and chromatin opening followed the gradual changes observed for 5hmC, whereas 5mC levels showed only mild fluctuations over time (Fig. 2A for up-DhMRs; Supplemental Fig. S2A for down-DhMRs). When comparing called peaks, 51.6% and 34.5% of the hydroxymethylated regions overlapped with H3K4me1 and FAIRE peaks, respectively, whereas 57% overlapped with H3K27ac peaks. These observations suggest a strong correlation between 5mC oxidation and both enhancer priming and activation during ECC differentiation. We next analyzed FOXA1, MEIS1, and PBX1 engagement at DhMRs and observed that, consistent with an enrichment in PF binding motifs at enhancers gaining 5hmC during differentiation (Sérandour et al. 2012), PF binding peaked at up-DhMRs (Fig. 2B) but not at down-DhMRs (Supplemental Fig. S2B), and heat maps clearly identified PF-bound regions as being enriched in 5hmC (Fig. 2C). In addition, these three PFs shared a significant fraction of their binding sites (Fig. 1D–G), indicating that FOXA1 and TALE-HD proteins target a common subset of enhancers. Collectively, these data suggest that the recruitment of these critical regulators at enhancers could influence or be sensitive to cytosine modifications.

Cytosine modifications increase TALE-HD binding in vitro

TALE-HDs bind DNA through the formation of MEIS1/PBX1 heterodimers or through dimerization with HOX proteins (for review, see Moens and Selleri 2006). ChIP-seq of MEIS1 and PBX1 in RA-treated P19 ECCs demonstrated that most MEIS1-bound regions are also interacting with PBX1 (Supplemental Fig. S3A). Accordingly, analysis of the proteins associated with PBX1-bound chromatin in RA-differentiated P19 cells by ChIP coupled to mass spectrometry (ChIP-MS) revealed the presence of several HOX and TALE-HD proteins, including MEIS1 (Supplemental Fig. S3B; Supplemental Table 1). In accordance with a putative pioneer function of TALE-HDs, both MEIS1 and PBX1 from NPC nuclear extracts can be efficiently captured by an artificial probe containing 5mCpGs in a TG(ACG)4 sequence (Spruijt et al. 2013). In addition, MEIS1 and PBX1 engage up-DhMRs subsequently to 5hmC occurrence during RA-induced P19 cell differentiation (Sérandour et al. 2012), suggesting that they could sense cytosine modifications rather than induce them. To address the possible influence of cytosine modifications on the binding of TALE-HD proteins to their recognition sites in vitro, pull-down analyses of cell extracts from differentiated P19 cells were then performed using probes containing the TGA(TA)C PBX1/HOXA9 motif (Shen et al. 1997) in which the cytosine from each strand of the CpG dinucleotide was differentially modified (either methylated or hydroxymethylated), as well as probes containing the MEIS1/ PBX1 motif TGA(TA)C (Chang et al. 1997) modified at underlined positions. For each probe, the specific binding of MEIS1/2 and PBX1/2/3 was first verified by supershift experiments (Supplemental Fig. S3C). Pull-down assays revealed that MEIS1 and PBX1/2/3/4 were better retained by the PBX1/HOXA9 probe containing specific combinations of modified cytosines (such as hemi-methylated, fully hydroxy- or methylated, or both methylated and hydroxymethylated probes) compared to the probe containing
Figure 1. Pioneer factors and NPC differentiation. (A) Pluripotent P19 EC cells derive from a 6.5-dpc mouse embryo and can differentiate into neural progenitor cells (NPCs) when treated with retinoic acid (RA). Pluripotency factors expressed in these cells include POU5F1, NANOG, SOX2, and NR5A1. Under the influence of RA, FOXA1, MEIS1, and PBX1 are induced and control the commitment of P19 EC cells to NPCs which express marker genes like Tshz1, Pax6, and Irx3. (B) RT-qPCR analysis of marker gene expression during 48 h of NPC differentiation after addition of RA. Curves were fitted with the GraphPad Prism software. (C) NPC marker genes are direct targets of TALE-HDs. Integrated Genome Browser (IGB) screenshots showing MEIS1 and PBX1 ChIP-seq signals at Irx3, Pax6, and Tshz1 loci in RA-treated cells. (D) IGB screenshot of FOXA1 and MEIS1 ChIP-seq signals from NPCs at the HoxB cluster, showing that FOXA1 and MEIS1 share common binding sites. (E) Heat maps showing FOXA1, MEIS1, and PBX1 ChIP-seq signals at 8278 FOXA1 sites in P19 cell-derived NPCs. (F) Logos of motifs enriched at FOXA1 binding sites. (G) Distribution of FOXA1 and PBX1 motifs in 50-bp windows centered on FOXA1 binding sites.
unmodified cytosines (Fig. 3A; Supplemental Fig. S3D). Likewise, MEIS1 and PBX1/2/3/4 proteins were better retained, although less markedly, by PBX1/MEIS1 probes containing modified cytosines (Fig. 3A). These results suggest that MEIS1 and PBX1/2/3/4 bind preferentially to sequences containing modified cytosines in vitro. This is consistent with data from Spruijt et al. (2013) indicating that MEIS1 and PBX1 are preferentially retained by fully methylated CpGs and extends their findings by showing that TALE-HDs also prefer CpGs where 5mC is combined with 5hmC on the other strand rather than to an unmodified cytosine.

We next evaluated the presence of 5hmCpGs at 20,287 MEIS1 binding sites in vivo by using single-CpG resolution data obtained by SCL-exo-seq in RA-treated P19 cells (Sérandour et al. 2016). This method allows the genome-wide identification of hydroxymethylated CpGs by a combination of chemical labeling (glucosylation followed by biotinylation), immobilization on streptavidin beads, and exonuclease digestion of the bead-trapped DNA molecules (Sérandour et al. 2016). As shown in Figure 3B, 5hmCpGs were enriched at bulk MEIS1 binding sites (see also Supplemental Fig. S4A). A cluster of 958 regions with centered 5hmCpGs was next recovered using the TMEV software (Saeed et al. 2003) in order to only consider regions with 5hmCpGs located within a few base pairs of MEIS1 binding sites. Sites from this cluster had, on average, a higher CpG density compared to bulk MEIS1 sites (Supplemental Fig. S4B). Heat maps, as well as profiles of MEIS1 ChIP-seq and SCL-exo signals (Fig. 3B–D), showed that 5hmCpGs accumulated within −15/+15 bp of the MEIS1 ChIP-seq peak summits (defined as 0 in the heat maps and the average profiles). The 958 sequences were next analyzed and a consensus sequence was defined with the CLC Sequence Viewer 6 software (Fig. 3D). Whereas 95 regions contained the TGACAG consensus MEIS1 site in close vicinity (within 2 to 20 bp) to 5hmCpGs, 29 regions did not include the consensus site but showed a TGACGG site, 24 of which had a hydroxymethylated CpG detected by SCL-exo (see examples in Supplemental Fig. S4C). De novo motif discovery from the 958 sequences was then implemented with both SeqPos (Fig. 3E; Supplemental Fig. S4D) and SEME (Supplemental Fig. S4E). SeqPos recovered a TGACG motif with a high probability, whereas the probability to find the MEIS1 motif was lower (Fig. 3E). The SEME tool ranked first the motif TGAC/TGGATGG (Supplemental Fig. S4E), which closely resembles the PBX1 motif TGATTGAT (Fig. 1F). In addition, despite being enriched in MEIS1 and PBX1 motifs, as shown by de novo motif discovery (Supplemental Fig. S4F), MEIS1 binding regions devoid of CpGs were characterized by a lower average MEIS1 ChIP-seq signal than the cluster of 958 MEIS1 sites with centered 5hmCpGs (Fig. 3F). Finally, the ability of MEIS1 to interact with a TGACGG sequence was assessed by competition pull-down experiments in which a fivefold excess of a

Figure 2. De novo hydroxymethylated enhancers bind pioneer factors driving NPC commitment. (A) Average kinetic profiles of 5hmC, 5mC, H3K4me1, H3K27ac, and chromatin opening (FAIRE) at 20,492 regions gaining 5hmC (up-DhMRs) at 48 h of P19 ECC differentiation. (B) Average profiles of FOXA1, PBX1, and MEIS1 binding at up-DhMRs at 48 h of differentiation. (C) Heat map representations of 5hmC signal at MEIS1, PBX1, and FOXA1 enriched regions in 48 h RA-treated P19 cells.

Maheé et al. 950 Genome Research www.genome.org

Downloaded from genome.cshlp.org on July 5, 2017 - Published by Cold Spring Harbor Laboratory Press
double-stranded oligonucleotide containing the TGACGG sequence was added to the PBX1/MEIS1 probe. Results indicated that indeed MEIS1 binds to the TGACGG site and that the fully methylated and fully hydroxymethylated sequences tend to show a stronger interaction with MEIS1 (Supplemental Fig. S4G). Collectively, these data suggest that MEIS1 can bind in vivo to TGACGG sequences including 5hmCpGs, although it might concern only a fraction of its target regions.

Cytosine modifications are required for the correct timing of neural progenitor differentiation

We next perturbed the DNA methylation and demethylation processes in P19 cells by using subtoxic concentrations of either 5-aza-2′-deoxycytidine (5-azadC), a hypomethylating agent which forms adducts with DNMTs (Jüttermann et al. 1994), or dimethoxyalgalicycine (DMOG), a cell-penetrating derivative of N-Oxalylglycine (NOG) that inhibits Fe2+/2-OG dioxygenases such as TET enzymes (Rose et al. 2011). As assessed by dot blot analysis, 5-azadC treatment reduced global 5mC levels in both undifferentiated and differentiated P19 cells and also limited 5hmC formation in RA-treated cells (Supplemental Fig. S5A). Treatment of P19 ECCs with DMOG did not impact global 5mC levels but inhibited significantly the increase in 5hmC observed during RA-induced differentiation (Supplemental Fig. S5B,C). We next evaluated the impact of the perturbation of 5mC and 5hmC levels on the production of mRNAs encoding pluripotency factors (POUSF1 and NANOgé), neural progenitor markers (TSHZ1, PAX6, and IRX3), and cell fate determination TFs (FOXA1, MEIS1, and PBX1) during RA-induced P19 cell differentiation (Supplemental Fig. S6A). Although RA-mediated changes in expression of pluripotency factors and Meis1 were not dramatically affected by 5-azadC and DMOG treatments, Foxa1 and Pbx1 mRNA levels did show some sensitivity to the drugs. However, such sensitivity was not observed at the protein level (Supplemental Fig. S6B). Notably, induction of Pax6 and Irx3 expression by RA was significantly delayed by 5-azadC and DMOG treatments (Supplemental Fig. S6A), suggesting that both DNMT and TET inhibition, although not affecting the repression of pluripotency factor genes, delayed differentiation of neural progenitors. This is consistent with the observed delayed induction of a subset of NPC marker genes during neural differentiation of Tet2−/− mouse ESCs (Hon et al. 2014). Since both RA-mediated loss of pluripotency and PF induction were not affected by the drugs while neural differentiation was delayed, these results indicate that our experimental conditions are suitable for the study of the role of DNA modifications at PF-bound enhancers.

Cytosine modifications control enhancer priming

In order to determine the role of cytosine modifications in enhancer priming and activation, we next analyzed the effects of 5-azadC or DMOG treatments on the levels of 5mC, 5hmC, H3K4me1, H3K27ac, and chromatin opening (FAIRE), as well as on MEIS1 and FOXA1 recruitment, at enhancers selected from genome-wide DIP- and ChIP-seq data according to (1) gain of 5hmC during differentiation, (2) enrichment in H3K4me1 and H3K27ac, and (3) recruitment of MEIS1 and/or FOXA1. Hence, 15 enhancers were selected, with five bound by MEIS1 but not by FOXA1 (enhancers M1 to M5), five recruiting both MEIS1 and FOXA1 (enhancers MF1 to MF5), five recruiting both MEIS1 and FOXA1 but not by MEIS1 (enhancers MM1 to MM5), and five recruiting only FOXA1 (enhancers FM1 to FM5). As assessed by dot blot analysis, 5-azadC treatment reduced global 5mC levels in both undifferentiated and differentiated P19 cells and also limited 5hmC formation in RA-treated cells (Supplemental Fig. S5A). Treatment of P19 ECCs with DMOG did not impact global 5mC levels but inhibited significantly the increase in 5hmC observed during RA-induced differentiation (Supplemental Fig. S5B,C). We next evaluated the impact of the perturbation of 5mC and 5hmC levels on the production of mRNAs encoding pluripotency factors (POUSF1 and NANOgé), neural progenitor markers (TSHZ1, PAX6, and IRX3), and cell fate determination TFs (FOXA1, MEIS1, and PBX1) during RA-induced P19 cell differentiation (Supplemental Fig. S6A). Although RA-mediated changes in expression of pluripotency factors and Meis1 were not dramatically affected by 5-azadC and DMOG treatments, Foxa1 and Pbx1 mRNA levels did show some sensitivity to the drugs. However, such sensitivity was not observed at the protein level (Supplemental Fig. S6B). Notably, induction of Pax6 and Irx3 expression by RA was significantly delayed by 5-azadC and DMOG treatments (Supplemental Fig. S6A), suggesting that both DNMT and TET inhibition, although not affecting the repression of pluripotency factor genes, delayed differentiation of neural progenitors. This is consistent with the observed delayed induction of a subset of NPC marker genes during neural differentiation of Tet2−/− mouse ESCs (Hon et al. 2014). Since both RA-mediated loss of pluripotency and PF induction were not affected by the drugs while neural differentiation was delayed, these results indicate that our experimental conditions are suitable for the study of the role of DNA modifications at PF-bound enhancers.

5hmC roles at enhancers

Figure 3. DNA binding by TALE-HD proteins is sensitive to cytosine modifications. (A) MEIS1 and PBX1/2/3/4 in vitro interaction with differentially modified synthetic PBX1/HOXA9 (left panel) and MEIS1/PBX1 (right panel) probes. Results are expressed according to the strongest binding equivalent to 100 (C: unmodified cytosine; M: methylated cytosine; H: hydroxymethylated cytosine; first letter: upper strand; second letter: lower strand). Statistical significance of differential binding to modified probes compared to the unmodified probe was tested with a Mann-Whitney test (n = 5; MEIS1 protein: [*] P < 0.05; [**] P < 0.01; PBX1/2/3/4: [#] P < 0.05, [##] P < 0.01). (B) Heat maps of MEIS1 ChIP-seq and SCL-exo (5hmC) signals at 20,287 bulk MEIS1 binding sites (upper panels) and 958 MEIS1 sites showing a centered SCL-exo signal (lower panels) in P19 cell-derived NPCs. (C) Average MEIS1 ChIP-seq (red curve) and SCL-exo (gray curve) signals at the 958 MEIS1 binding sites showing a centered SCL-exo signal in P19 cell-derived NPCs. Note that oscillations in the MEIS1 profile are due to the lower resolution of the MEIS1 ChIP-seq .wig file compared to the SCL-exo .wig file. (D) Average SCL-exo signal over 40 bp for the cluster of 958 MEIS1 sites depicted in B. The consensus sequence determined from the aligned 958 sequences with the software CLC Sequence Viewer 6 is indicated on the graph. A motif related to the TGACAG consensus MEIS1 motif has been highlighted in red font. (E) Examples of de novo motifs found enriched (P-value is indicated above the logos) by the SeqPos motif tool from Cistrome. (F) Box plot analysis of MEIS1 ChIP-seq signal for MEIS1 sites with no CpGs within the ChIP-seq peaks (n = 3628) and the 958 MEIS1 sites with centered SCL-exo signal (m = mean ChIP-seq signal).
to MF5), and five bound by FOXA1 but not by MEIS1 (enhancers F1 to F5). As a control, we selected an enhancer enriched in H3K4me1 and H3K27ac and bound by both FOXA1 and MEIS1 but having no CpG in 400 bp flanking FOXA1 and MEIS1 ChIP-seq peak summits (enhancer “No CpG”). Notably, the chromatin features of this control enhancer were not affected by 5-azadC or DMOG treatments (Supplemental Fig. S7A,B). Since DMOG could also inhibit histone demethylases of the Jumonji family (Rose et al. 2011), we controlled whether DMOG treatment increased H3K27 methylation, potentially leading to inhibition of enhancers, and found that H3K27me3 levels did not increase in DMOG-treated cells for any of the selected enhancers (Supplemental Fig. S7C).

These data indicate that, in our conditions, the observed effects of the DMOG treatment mainly reflected TET inhibition. The chromatin features of these 15 enhancers were then analyzed by (h)MeDIP-, ChIP-, and FAIRE-qPCR experiments. Results of these experiments are depicted in the histograms of Supplemental Figures. S8 and S9 and were analyzed by principal component analysis (PCA) with data expressed as fold change between control cells and cells treated with the inhibitor, with a scale of 0.5 for 5-azadC (A,C) and 0.25 for DMOG (B,D). Variations in MEIS1 and FOXA1 recruitment are represented below heat maps in panels C and D but were not used for hierarchical clustering.

Figure 4. Cytosine modifications regulate enhancer priming. Dendrograms and heat maps from hierarchical clustering of enhancer features (left dendrogram) and the 15 selected enhancers studied (top dendrogram). (A,B): undifferentiated P19 ECCs. (C,D): differentiated neural progenitors. The color scale bars at the bottom of each column indicate log2-transformed fold-change between control cells and cells treated with the inhibitor, with a scale of 0.5 for 5-azadC (A,C) and 0.25 for DMOG (B,D). Variations in MEIS1 and FOXA1 recruitment are represented below heat maps in panels C and D but were not used for hierarchical clustering.

Whenever present in the selected enhancers, taking advantage of the fact that MspI cannot cut a C(5hm)CGG site after glucosylation of the ShmC (eight out of 15 enhancers) (Supplemental Fig. S12). Results obtained with undifferentiated P19 cells showed that inhibition of both DNMTs (Fig. 4A) and TETs (Fig. 4B) mainly affected ShmC and H3K4me1 levels. Remarkably, under DNMT inhibition, HCPC identified two clusters of enhancers: one characterized by a decrease in ShmC at sites having significant levels of ShmC in control condition (Supplemental Fig. S13: MF1, F4, M5, and F2), consistent with a decrease in the substrate for TET-mediated oxidation; and a second one grouping regions with increased levels of H3K4me1, indicating that 5mC prevents mono- methylation of H3K4 (Fig. 4A). In addition, enhancers showing a decrease in ShmC in the presence of DMOG also showed a decrease in H3K4me1 levels in undifferentiated cells (Fig. 4B). This is consistent with data showing that, in MCF-7 cells, binding of the H3K4 methyltransferase KMT2A (also known as MLL1) to the TFF1 enhancer depends on DNA hypomethylation (Jeong et al. 2014) and that 5mC repels KMT2A in vitro (Spruijt et al. 2013).

Importantly, data obtained with differentiated P19 cells revealed a positive correlation between the levels of ShmC and chromatin opening (Fig. 4C,D). Since inhibition of 5mC formation by 5-azadC did not impact chromatin opening (FAIRE) in undifferentiated cells despite an increase in H3K4me1 levels (Fig. 4A),
that 5hmC and TET levels were inversely correlated for three clusters out of four (Fig. 5), suggesting that high TET residency may indeed be associated with high 5hmC superoxidation. Hence, FOXA1 binding to enhancers might favor the engagement of TET enzymes involved in 5hmC oxidation. Accordingly, it was recently demonstrated that FOXA1 interacts with TET2 in prostate cancer cells (Takayama et al. 2015) and that 5hmC oxidation in human ECCs is mainly dependent on TET2 (Putiri et al. 2014). Similarly, MEIS1 binding sites were clustered according to their 5hmC and MEIS1 levels, and the clusters were analyzed for TET1 and TET2 engagement (Supplemental Fig. S16). Clusters C2 and C3 presented low and high 5hmC levels, respectively, and were similarly associated with TET1, whereas C2 was enriched in TET2 and, interestingly, presented the highest enrichment in FOXA1. These data, together with results from DMOG and siRNA experiments, strongly suggest that FOXA1-only and FOXA1/MEIS1-bound enhancers are more likely to undergo 5hmC superoxidation through TET2 recruitment than enhancers bound only by MEIS1.

Although MEIS1 engagement was decreased at eight out of the 10 MEIS1-bound enhancers (statistically significant only for four of them) in the 5-azadC condition (Fig. 4C), the impact of DMOG treatment on its recruitment was variable but had a tendency to follow changes in chromatin opening levels at seven out of 10 enhancers (Fig. 4D). These observations are consistent with a model (Supplemental Fig. S17) in which MEIS1 binding could be stabilized by 5mC/5hmC (in agreement with in vitro data shown in Fig. 3A) and facilitated by chromatin opening. Accordingly, the rapid increase in 5hmC levels at MEIS1 binding regions was systematically followed by chromatin opening and MEIS1 recruitment (Fig. 6), and a significant fraction of these sites (Fig. 6, clusters 1–2, including 1032 regions out of 3366) showed chromatin opening before MEIS1 engagement. In turn, binding of MEIS1 may favor acetylation of H3K27, as suggested by the correlation between these two events (Fig. 4; Supplemental Figs. S8, S9). All together, these results indicate that 5mC removal is required for enhancer priming and that 5hmC per se is functionally linked to chromatin opening events allowing recruitment of MEIS1.

### Discussion

Whereas FOXA1 is known to bind nucleosomes (Sekiya et al. 2009), the data presented herein suggest that MEIS1 engagement at enhancers requires chromatin opening in correlation with TET-dependent 5mC oxidation. MEIS1-bound enhancers are particularly enriched in 5hmC, and we show evidence that MEIS1 can bind to sites including 5hmCpGs. Hence, 5mC oxidation could favor MEIS1 binding to DNA by promoting chromatin opening, suggesting that MEIS1 might poorly bind to nucleosomal DNA. Interestingly, the related TALE-HD protein PBX1, which dimerizes with MEIS1 in P19-derived NPCs, is enriched in pull-down experiments with naked methylated DNA but not with methylated DNA wrapped around histones (Bartke et al. 2010). These in vitro experiments allow recruitment of MEIS1.

**Figure 5.** High levels of 5hmC correlate with low levels of TETs at FOXA1-bound enhancers. (A) Heat maps of FOXA1 and 5hmC signals in clustered FOXA1 binding regions outside CpG islands. (B) Average profiles of 5hmC, TET1, and TET2 recruitment at the C2, C3, C4, and C5 clusters shown in A.
observations are in agreement with a model in which 5hmC first favors nucleosome instability in vivo and then stabilizes PBX1/MEIS1 on DNA at CpG-containing recognition sites. Nonetheless, we cannot exclude that MEIS1 binds actually to sites with unmodified cytosines after complete DNA demethylation. Oxidative and classical hairpin bisulfite sequencing of DNA recovered from MEIS1-bound immunoprecipitated chromatin could eventually bring the answer to this question, but purifying the required amount of ChiPed DNA appears challenging.

By inhibiting DNA methylation, we observed a remarkable increase in H3K4me1 at a majority of tested enhancers, strongly suggesting that 5mC is a repellent for H3K4 methyltransferases in vivo. In accordance with this idea, KMT proteins are directly repelled by 5mC in vitro (Birke et al. 2002; Bach et al. 2009; Spruijt et al. 2013), suggesting that oxidation of 5mC could be sufficient for allowing KMT recruitment to CpG-containing DNA sequences through their CXXC domain and methylation of H3K4. In agreement with this hypothesis, our kinetic ChiP-seq data as well as the PCA analysis of DMOG-treated P19 E1Cs indicated a strong correlation between demethylation of H3K4 and the presence of 5hmC. Interestingly, the recently released structure of the human KMT2A CXXC domain in complex with CpG-containing DNA (PDB, 4NW3) reveals that the CXXC domain is anchored to both cytosines of the CpG by two residues, namely K1186 and Q1187, which establish contacts with the fifth carbon of the cytosines. Hence, the presence of a methyl group or a hydroxymethyl group on the fifth carbon would be predicted to destabilize the interaction between the CXXC domain of KMT2A and DNA, as observed in vitro by Spruijt et al. (2013). Accordingly, sites which gain 5hmC in the presence of DMOG (most likely because of a reduced oxidation of 5mC), and thus which have lower occurrence of a complete DNA demethylation process, should have reduced levels of H3K4me1. Our data actually show the opposite, since four out of the five enhancers gaining H3K4me1 in DMOG + RA treated cells also gained 5hmC (Fig. 4D), suggesting that the hydroxymethyl group might favor additional interactions with the CXXC domain of KMT2A.

Conversely to what was observed for H3K4me1, our data indicate that the presence of 5hmC itself at enhancers could facilitate chromatin opening, leading to the hypothesis that this modified base plays a role in the destabilization or eviction of nucleosomes. Still, it cannot be excluded that chromatin opening could occur through the action of remodelers or histone chaperones that would be brought to enhancers by TET enzymes and would not require erasure of cytosine modifications per se. In that case, 5hmC would merely reflect the presence of a TET enzyme and would not, stricto sensu, participate in chromatin opening. If this hypothesis was correct, then blocking superoxidation with DMOG should not lead to an increase in FAIRE signal. Nonetheless, such an increase was noted for five out of the seven enhancers that experienced a gain in 5hmC in DMOG + RA-treated cells, indicating that 5hmC might indeed be directly involved in the regulation of chromatin opening. Interestingly, apart from PBX1, PBX2, MEIS1, MEIS2, and PKNOX2 TALE-HD proteins, and their HOX partners, we identified PBX1-bound chromatin as being enriched in components of the chromatin remodeling complexes NuRD and FACT (Supplemental Table 1). NuRD is known to be associated with both active and repressed chromatin and can be targeted by MECP2, MBD2, and MBD3 CpG-binding proteins (Whyte et al. 2012; Günther et al. 2013; Shimbo et al. 2013). Whereas neither MECP2 nor MBD2 were identified in the PBX1 ChiP-MS experiment, MBD3 was (Supplemental Table 1). These data are consistent with the observed engagement of NuRD and MBD3 at 5hmC-enriched chromatin regions (Yildirim et al. 2011; Baubec et al. 2013). Nonetheless, MBD3 is still found at these regions in Dnmt3 triple-KO mouse ES cells, suggesting that 5hmC does not play a direct role in MBD3 targeting (Baubec et al. 2013). The presence of FACT at PBX1-bound enhancers indicates that the chromatin specificities of these sites could also impact FACT engagement. FACT facilitates H2A/H2B eviction from nucleosomes (Dinant et al. 2013), a process which relies on its ability to lower H2B interaction with DNA (Hondele et al. 2013). Interestingly, FACT activity can be enhanced by poly-ADP ribosylation of its subunit SPT16 by TET enzymes and would not, stricto sensu, participate in chromatin opening. If this hypothesis was correct, then block-
a recent study described the engagement of FACT at a POU5F1-bound enhancer in ES cells and its role in nucleosome eviction (Shakya et al. 2015). Interestingly, in this study, DMOG reduces the engagement of FACT at this POU5F1-bound enhancer, suggesting that FACT recruitment depends either on histone demethylation, or on DNA hydroxymethylation, or on both (Shakya et al. 2015).

Collectively, our data further validate the concept that DNA demethylation is required for the proper regulation of enhancer priming and activation during cell differentiation and provide additional clues for understanding how nonpromoter DNA (hydroxymethyl)ation can participate in the cell-specific control of gene activity.

Methods

Cell culture

P19.6 mouse embryonal carcinoma cells were grown and differentiated as described (Sérandour et al. 2012). To inhibit DNA methyltransferases, cells were plated in a medium containing 20 nM of 5-aza-2-deoxycytidine (5-azaDC, Sigma) for 24 h. The medium was then renewed and cells were cultured in the presence of 20 nM 5-azaDC for 48 h with or without RA treatment. Inhibition of TET proteins was performed by treating P19 cells with 0.5 mM of dimethylxalylglycine (DMOG, Cayman Chemical). Due to DMOG instability, the medium was renewed every 24 h. Tet1-, 2-, and 3-specific 27mer siRNA duplexes (Trilencer-27 siRNA) and universal negative siRNA duplex were purchased from Origene Technologies. Transfections of P19 cells with siRNAs at a final concentration of 10 nM were performed in Opti-MEM medium using Oligofectamine (Invitrogen) as recommended by the manufacturer. After 6 h of transfection, the medium was renewed and cells were cultured for 48 h with or without RA treatment.

Chromatin and DNA immunoprecipitation

ChIP experiments were carried out as previously described (Sérandour et al. 2011). ChIP followed by high-throughput sequencing (ChIP-seq) experiments were performed as in Schmidt et al. (2009). DNA immunoprecipitation protocols are available on the EpiGeneSys website (http://www.epigenesys.eu/en/protocols): “Methylated DNA Immunoprecipitation (MeDIP)” (Protocol 33 - Michael Weber and Dirk Schübele) and “Hydroxymethylated DNA Immunoprecipitation (hMeDIP) in mammalian cells” (Protocol 64 - Elise Mahé, Gilles Salbert).

Formaldehyde assisted isolation of regulatory elements

Formaldehyde assisted isolation of regulatory elements was performed as previously described (Eckhoute et al. 2009).

Sequencing and bioinformatics

The (h)MeDIP, ChIP, FAIRE, and input libraries were prepared using the TrueSeq ChIP Sample Prep kit (ref. IP-202-1012, Illumina). The libraries were sequenced by HiSeq following the manufacturer’s protocol. Alignment of reads, conversion of extracted reads to .wig signal files, and peak calling were performed as previously described (Sérandour et al. 2012). Blacklisted regions available from UCSC (http://genome.ucsc.edu) were removed from each .wig signal file. CpG island (CGI) coordinates were also obtained from UCSC. Signals were normalized in order to compare (h)MeDIP, FAIRE, or ChIP-seq signals from different time points in time-course experiments. Proper normalization requires both calibrating the highest peaks and accounting for noise heterogeneity. For MEIS1, which is an induced transcription factor, signals within .wig files were calibrated according to the number of mapped reads. For other signals, assuming that the highest peaks remained, on average, at similar levels, we selected 15,000 positions with the highest signal for each time point sample, determined their mean signal value, and chose as a reference the sample with the highest mean signal. Then, all signals within a given .wig file were calibrated by multiplication by a factor to \( a = s_{\text{ref}}/s_i \), where \( s_{\text{ref}} \) is the mean value of the reference signal and \( s_i \) the average value of signal \( s_i \) in their respective top 15,000 positions. In order to reduce artifacts linked to noise heterogeneity between samples, we first defined for each sample an upper noise threshold \( r \) as being twice the mean value of the signal for all positions of the corresponding .wig file. A peak-calling algorithm (Sérandour et al. 2012) was next used to identify peaks with a signal higher than \( r \) in at least three consecutive positions. Called peaks were then extended by 100 bp on each side to include positions corresponding to the sides of the peaks even though they had an initial signal below \( r \). Finally, the original signal present within these extended regions was retrieved to generate normalized noise-free .wig files. Heat maps were generated by the Cistrome web-platform (http://cistrome.org/ap/; Liu et al. 2011). Motif search was run online through the SeqPos tool from Cistrome, and the analysis of motif distribution in genomic regions was done using CentDist (http://biogps.ddns.comp.nus.edu.sg). TET1 ChIP-seq and SCI-exo data are from the NCBI GEO GSM941681 and GSE70635 data sets.

Pull-down assay

Differentially modified 20-bp synthetic oligonucleotides (Sigma), biotinylated on the forward strand and containing either the PBX1/HOXA9 (TGATTTAGC) or the PBX1/MEIS1 (TGATTGAC AG) recognition sites, were used as probes for pull-down assays. For each probe, cytosines from each CR dinucleotide (R = A or G) were unmodified (C), methylated (M), or hydroxymethylated (H). DNA/protein complexes were purified with streptavidin-coupled magnetic beads (Invitrogen) after incubation in binding buffer (Hepes 20 mM pH 7.9, 50 mM KCl, 1 mM DTT, 2.5 mM MgCl2, 5 µg/µL BSA, 10% glycerol) with 10, 20, or 40 µg of extracts from P19 cells treated for 48 h with RA. Protein binding was then analyzed by Western blot. Signals were quantified using the Quantity One 1-D Analysis Software from Bio-Rad. Competition assays were run with the biotinylated PBX1/MEIS1 probe and a fivefold excess of differentially modified double-stranded oligonucleotides of the following sequence: TACACTGATTGACGGTGCGAT.

ChIP-MS

ChIP followed by mass spectrometry analysis (ChIP-MS) was performed in triplicate as described in Mohammed et al. (2013). Samples were analyzed by the Proteomics core facility of the Cambridge Institute (Cancer Research UK). The nuclear fraction from 5 × 10⁷ P19 cells cross-linked with formaldehyde was extracted and sonicated. Immunoprecipitation was performed with 100 µL of magnetic beads (Dynal) prebound either with 20 µg of PBX1 antibody or with 20 µg of control IgG. After several washes of the beads, the bead-bound proteins were digested with trypsin. The beads were then incubated overnight at 37°C to allow the salvage of the digested peptides, which were then analyzed by nano-LC-MS/MS. Mass spectrometry was performed using an LTQ Velos-Orbitrap mass spectrometer (Thermo Scientific) coupled to an Ultimate RSLCnano-LC system (Dionex). Results were processed with the Scaffold Viewer software with protein threshold and peptide threshold set at 90%. The mass spectrometry proteomics data

www.genome.org
have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al. 2016) partner repository with the data set identifier PXD006158.

**Chemiluminescent EMSA**

EMSA experiments were performed with nuclear extracts (NEs) prepared from P19 cells treated for 48 h with 1 µM RA. Two micrograms of NE were incubated with binding buffer (Hepes 20 mM pH 7.9, 10 mM KCl, 1 mM DTT, 2.5 mM MgCl₂, 10% glycerol), 1 µg poly(dl-dC), and 150 fmol of the biotinylated probes used in pull-down assays. For supershift assays, NEs were first pre-incubated with the antibody of interest. EMSAs were performed using the “GelshiftTM Chemiluminescent EMSA” kit (Active Motif) following the manufacturer’s instructions.

**Dot blot assay**

Genomic DNA was extracted using a DNA extraction kit (DNeasy Blood and Tissue kit; Qiagen) with a supplementary step of RNase digestion. Three hundred nanograms, 150 ng, and 75 ng of each sample DNA was spotted to a nylon membrane (Hybond N, GE Healthcare) previously soaked in 2x SSPE solution (0.3 M NaCl, 20 mM sodium phosphate, 2 mM EDTA) and inserted in the dot blot apparatus (SCIE-PLAS). DNA was cross-linked to the membrane by 15 min exposure to UV light before blotting. After revelation, membranes were stained with methylene blue to control the amount of DNA spotted. Signals were quantified using the Quantity One 1-D Analysis Software from Bio-Rad.

**RNA preparation and reverse transcription**

RNA was isolated by TRIzol extraction (Invitrogen) and reverse-transcribed using M-MLV reverse transcriptase (Invitrogen) and Pd(N)6 random hexamers (Amersham Pharmacia Biosciences).

**Epimark 5hmC and 5mC analysis**

Methylation or hydroxymethylation status of selected CpGs included in a MspI CCGG restriction site was verified using the Epimark 5hmC and 5mC Analysis kit (New England Biolabs) according to the manufacturer’s protocol. Briefly, genomic DNA was first treated by T4 beta-glucosyltransferase (T4-BGT) in order to add a glucose moiety to the hydroxyl group of 5hmC, and then digested by either MspI or HpaII restriction enzymes. Both enzymes recognize a CCGG sequence, but HpaII cleaves only unmodified sites, whereas MspI cleaves 5mC or 5hmC but not 5ghmC. Approximate quantification of cytosine modifications was obtained by real-time PCR. Specific oligonucleotides were designed whenever those designed for ChIP-qPCR were not appropriate. Analysis and calculations were performed according to the kit protocol.

**Principal component analysis and clustering**

The R software (R Core Team 2013), with the FactoMineR package (Lé et al. 2008), was used to perform principal component analysis and hierarchical clustering. Briefly, chromatin mark variations across several genomic regions were investigated by PCA in which variables were genomic regions and individuals were chromatin marks. Data were sequentially scaled to unit variance and then projected onto orthogonal components. The resulting eigenvalues (based on both the covariance and correlation matrices) were used to retain the first components accounting for over 90% of the variance in the data. Finally, the coordinates of the individuals and variables on the selected principal components were used for the hierarchical classification.

**Real-time PCR (qPCR)**

Real-time PCR was performed using SYBR Green Master Mix (Bio-Rad, France) on a Bio-Rad CFX96 machine with 39 cycles of amplification. All primers were purchased from Sigma and are listed in the Supplemental Materials and Methods section.

**Statistical analyses**

Statistical analyses were performed using GraphPad Prism software. Statistical significance was determined using the Mann-Whitney test.

**Data access**

All sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) database under accession number GSE82314. PBX1 ChIP-MS data are accessible at the PRIDE Archive database from the European Bioinformatics Institute (https://www.ebi.ac.uk/pride/) under accession number PXD006158.

**Acknowledgments**

We thank C. Le Péron and A. Laurent for critical reading of the manuscript. We also thank the “Human and Environmental Genomics” platform (Rennes) for Illumina sequencing. We are thankful to C. D’Santos and Chris Taylor from the Proteomics Core Facility of the Cambridge Institute (Cancer Research UK) for running the MS experiments. This work was supported by the CNRS, the University of Rennes 1, and the Association de la Recherche contre le Cancer. E.A.M. was supported by a PhD fellowship from the Ministère de l’Enseignement Supérieur et de la Recherche.

**Author contributions:** E.A.M. planned and ran experiments, analyzed the data, and wrote the manuscript. T.M., A.A.S., M.B., and G.P. ran experiments. S.A. ran bioinformatic analyses. F.C. designed and ran ChIP-qPCR experiments and analyzed the data. G.S. planned the study, analyzed the data, and wrote the manuscript.

**References**


Received June 16, 2016; accepted in revised form April 5, 2017.