BMC
Genomics

RESEARCH ARTICLE

Open Access

5-hydroxymethylcytosine represses the activity of enhancers in embryonic stem cells: a new epigenetic signature for gene regulation

Inchan Choi[†], Rinho Kim[†], Hee-Woong Lim, Klaus H Kaestner[*] and Kyoung-Jae Won[*]

Abstract

Background: Recent mapping of 5-hydroxymethylcytosine (5hmC) provides a genome-wide view of the distribution of this important chromatin mark. However, the role of 5hmC in specific regulatory regions is not clear, especially at enhancers.

Results: We found a group of distal transcription factor binding sites highly enriched for 5-hdroxymethylcytosine (5hmC), but lacking any known activating histone marks and being depleted for nascent transcripts, suggesting a repressive role for 5hmC in mouse embryonic stem cells (mESCs). 5-formylcytosine (5fC), which is known to mark poised enhancers where H3K4me1 is enriched, is also observed at these sites. Furthermore, the 5hmC levels were inversely correlated with RNA polymerase II (PolII) occupancy in mESCs as well as in fully differentiated adipocytes. Interestingly, activating H3K4me1/2 histone marks were enriched at these sites when the associated genes become activated following lineage specification. These putative enhancers were shown to be functional in embryonic stem cells when unmethylated. Together, these data suggest that 5hmC suppresses the activity of this group of enhancers, which we termed "silenced enhancers".

Conclusions: Our findings indicate that 5hmC has a repressive role at specific proximal and distal regulatory regions in mESCs, and suggest that 5hmC is a new epigenetic mark for silenced enhancers.

Keywords: 5hmC, GROseq, PolII, eRNA, mESC, Enhancer

Background

5-hydroxymethylcytosine (5hmC) is an epigenetic mark that arises from oxidation of 5-methylcytosine (5mC) by Ten-eleven translocation (Tet) enzymes [1,2]. The 5hmC mark has been studied in several cell types, such as mouse embryonic stem cells (mESCs) [2,3], neuronal cells [4-6] and adipocytes [7]. 5hmC is enriched at promoters marked bivalently by H3K4me3 and H3K27me3 in mESCs [8], but depleted at promoters in the brain [9]. 5hmC is also enriched at specific transcription factor binding sites (TFBSs) in human and mouse ESCs [1,9-14]. Specifically, in mESCs, 5hmC is depleted at Sox2 and Oct4 binding sites, but enriched for Esrrb and Tcfcp2l1 occupancy [12].

In human embryonic stem cells (hESCs), 5hmC is highly enriched at CTCF, Nanog, and Oct4 binding sites [11]. Another study in hESC observed that the 5hmC profile showed a bimodal distribution at Oct4, Sox2, TAF1 and p300 binding sites [9]. While these studies suggest a possible regulatory role for 5hmC at promoters and TFBSs, its function at these regulatory regions remains unclear.

Here, we report on a new repressive role for 5hmC at specific regulatory regions in mESCs. We show that 5hmC negatively correlates with nascent transcripts, especially at TFBSs. Interestingly, we discovered that a group of distal TFBSs displays a new epigenetic signature; these sites are exclusively enriched for 5hmC, depleted for activating histone modification marks (H3K4me1 and H3K27ac), and significantly reduced for nascent transcripts or enhancer RNAs (eRNAs). The expression of the genes close to these TFBSs was significantly lower than that of genes close to other classes of TFBSs. In addition, we found that a fraction of these TFBSs becomes enriched for activating

* Correspondence: kaestner@mail.med.upenn.edu; wonk@mail.med.upenn.edu
[†]Equal contributors
Department of Genetics, Institute for Diabetes, Obesity and Metabolism, University of Pennsylvania, 3400 Civic Center Blvd, 19104 Philadelphia, PA, USA

BioMed Central

histone marks (H3K4me1/2) in neural progenitor cells (NPCs) or endomesoderm cells. RNA polymerase II (PolII) chromatin interaction analysis with paired-end tagging (ChIA-PET) [15] showed that the target genes of these regulatory regions were indeed significantly up-regulated in NPCs. Enhancer/luciferase reporter assays demonstrated that these regions function as in gene activation when 5hmC is removed for these sites. Together, our findings suggest that 5hmC is as a novel marker for transcriptional silent enhancers in mESCs for regulatory regions that are activated during development.

Results

A group of 5hmC-enriched distal TFBSs is lacking activating histone marks and nascent RNA transcription
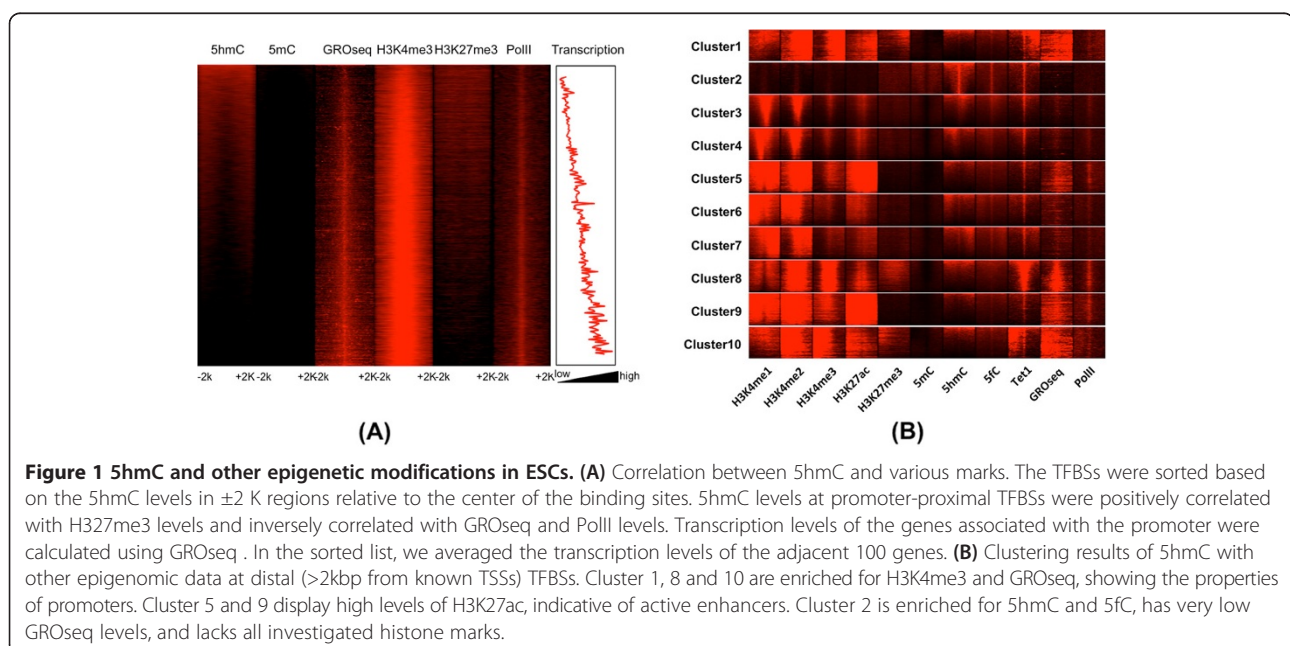
A recent survey had found 5hmC enriched at TFBSs in hESC [11], mouse neuronal cells, and adipocytes [7]. Therefore, we investigated 5hmC levels [13] at the binding sites of 13 key transcription factors (TFs) (Nanog, Oct4, STAT3, Smad1, Sox2, Zfx, c-Myc, n-Myc, Klf4, Esrrb, Tcfcp2l1, E2f1 and CTCF) in mESC [16]. We confirmed previous results [11,12] that 5hmC was generally depleted at the core of the proximal (within 2 kb to transcription start sites (TSSs)) TFBSs, but relatively high in the regions neighboring (±2 kb) the core (Additional file 1: Figure S1A). We also confirmed that 5hmC is highly enriched at the core of distal binding sites of many TFs, such as Zfx and Esrrb (Additional file 1: Figure S1B) [11,12].

To further investigate the role of 5hmC in gene regulation in conjunction with other epigenetic marks, we performed an integrative analysis using 5hmC, 5mC [13], Tet1 [10], H3K4me1/2/3, H3K27me3, RNA polymerase (Pol) II occupancy [17] and nascent RNAs from global run-on sequencing (GROseq) [18] data. We found that 5hmC levels were inversely correlated with nascent RNA transcription and Pol II occupancy at proximal TFBSs (Figure 1). We confirmed the levels of 5hmC positively correlated with the levels of the repressive H3K27me3 histone mark at proximal TFBSs [8,12].

To study the epigenetic landscapes surrounding distal TFBSs, we applied the K-means algorithm (K = 10) and found clusters marked by various epigenetic modifications (Figure 1B). Clusters 1, 8 and 10 showed the properties of active promoters: H3K4me2/3 enrichment with relatively low levels of H3K4me1 and the presence of nascent RNA transcripts. These clusters thus likely represent the promoters of long intergenic non-coding RNAs [19] or unannotated promoters of protein-coding genes. Clusters 5 and 9 showed H3K4me1 and H3K27ac enrichment, indicating active enhancers. These clusters, as well as clusters 3, 4, 6, and 7, showed only a small amount of nascent transcripts or enhancer RNAs (eRNAs), which have been known to correlate with the gene transcription levels of adjacent genes [20,21]. The presence of eRNAs in these clusters suggest that the TFBS at these clusters have an activating role.

We were especially interested in cluster 2, which was enriched for 5hmC, but was depleted of eRNAs. Strikingly, this cluster had no activating histone marks such as H3K4me1 or H3K27ac [22-24], even though TFs bind at these sites (Figure 1B and Additional file 1: Figure S2). 5mC was depleted at the core of the TFBS, consistent with the previous observation in hESCs [25]. Compared with other clusters, cluster 2 was characterized by low levels of

Figure 1 5hmC and other epigenetic modifications in ESCs. (A) Correlation between 5hmC and various marks. The TFBSs were sorted based on the 5hmC levels in ±2 K regions relative to the center of the binding sites. 5hmC levels at promoter-proximal TFBSs were positively correlated with H327me3 levels and inversely correlated with GROseq and PolII levels. Transcription levels of the genes associated with the promoter were calculated using GROseq . In the sorted list, we averaged the transcription levels of the adjacent 100 genes. **(B)** Clustering results of 5hmC with other epigenomic data at distal (>2kbp from known TSSs) TFBSs. Cluster 1, 8 and 10 are enriched for H3K4me3 and GROseq, showing the properties of promoters. Cluster 5 and 9 display high levels of H3K27ac, indicative of active enhancers. Cluster 2 is enriched for 5hmC and 5fC, has very low GROseq levels, and lacks all investigated histone marks.

eRNAs and low PolII occupancy. To confirm the enrichment for 5hmC, we investigated the profile of sequencing data from other independent studies [1,12-14,26,27]. Cluster 2 was enriched for 5hmC consistently for all four independently measured datasets (Additional file 1: Figure S3). We also examined TAB-seq, which provides base-resolution sequencing of 5hmC in mESC [3]. The TAB-seq profile also confirmed enrichment for 5hmC at the core of TFBSs for cluster 2 regions for both strands (Additional file 1: Figure S4). Together, these data suggest that 5hmC combined with absence of H3K4me1 at distal TFBSs marks inactive enhancers.

Surprisingly, cluster 2 is also highly enriched for 5-formylcytosine (5fC) compared with other clusters (Figure 1B). Both 5fC and 5hmC are involved in the active demethylation pathway [28,29]. Previous genome-wide study using 5fC revealed that 5fC is enriched at enhancers, especially at poised enhancers marked by H3K4me1 without H3K27ac [30]. However, the properties of the cluster 2 regions are novel, as they lack the H3K4me1 mark. This strongly suggests that 5hmC as well as 5fC mark a novel type of "poised" or silenced enhancer at distal regulatory regions where active histone modification marks are absent.

Next, we interrogated the state of the 5hmC mark in other cell types. In hESCs, we also identified a cluster enriched for 5hmC [3] but depleted for both H3K4me1 and H3K27ac at distal DNaseI hypersensitive sites (DHSs) [31] (Additional file 1: Figure S5). As in mESCs, GROseq levels in hESCs [32] were significantly weaker in this cluster (p-value = 1.7e-14). In mature adipocytes, we observed 5hmC [7] enriched at over 20% of PPARγ binding sites [33] (Additional file 1: Figure S6). Surprisingly, PolII occupancy [33] was depleted when 5hmC was enriched (Additional file 1: Figure S6). These data indicate that 5hmC can be a repressive mark at distal regulatory regions regardless of cell type or differentiation state.

Additional file 1: Table S1 lists the number of binding sites for each TF in cluster 2 in mESCs. The majority of the cluster 2 regions were bound by CTCF, Tcfcp2l1 or Esrrb. Fewer binding sites for Oct4, Sox2, and Nanog, the master regulators for self-renewal and pluripotency in ESCs, were observed in cluster 2 [34]. This is consistent with the observation that 5hmC is depleted at highly active enhancers in ESCs. We further investigated if ChIP intensity is lower for the TFBSs in cluster 2. We did not find statistical differences, even though the average profiles of the TFBSs in cluster 2 were slightly lower compared with the TFBSs in other clusters (Additional file 1: Figure S7).

5hmC-enriched distal TFBSs are associated with developmental genes

Next, we analyzed the correlation between 5hmC levels and transcriptional activity of the genes closest to the TFBSs for each cluster. To calculate gene transcription levels, we calculated the reads per kilobase per million mapped reads (RPKM) from GROseq (see Methods). The genes mapping to the TFBSs in cluster 2 had strikingly reduced transcription levels compared to the genes in all other clusters (p-value <1.3e-20), even compared to clusters 8 and 10, where the repressive H3K27me3 mark was relatively enriched (Figure 1B).

GO analysis of the genes closest to the TFBSs in cluster 2 using GREAT [35] revealed that the genes in this cluster were enriched for developmental functions, such as "muscle cell development" (p-value = 3.4e-14)" and "foregut morphogenesis" (p-value = 5.8e-9) (Figure 2D). This is consistent with the fact that these genes are silent in ESCs and are only activated once differentiation commences.

A snapshot in Figure 3 shows the enrichment for 5hmC at the Klf4 and the Esrrb binding sites located in the first intron of *Sorcs2. Sorcs2* is highly expressed in the developing and mature murine central nervous system [36]. We observed that *Sorcs2* is silent in mESC, and its promoter is bivalently marked by H3K4me3 and H3K27me3 [17]. In mouse neural progenitor cells (NPCs), however, *Sorcs2* is highly expressed [17]. The Klf4 and the Esrrb binding sites are marked by H3K4me1 in NPCs, suggesting an active role of this region as an enhancer during neural development.

5hmC-enriched distal TFBSs become activated during development in a lineage-specific way

Because of the coordination of high 5hmC levels with low expression of genes in cluster 2, we hypothesized that 5hmC may attenuate enhancer activity in mESCs, which becomes activated later during development. Therefore, we analyzed H3K4me1/2 data in NPCs [17] and endomesoderm cells [37]. H3K4me1 and H3K4me2 are known to mark enhancers [22]. While H3K4me1/2 enhancer marks were depleted in mESCs, around 9% of distal TFBSs (out of 5,278 TFBSs) showed enriched H3K4me1/2 occupancy in NPCs, and an additional 20% of TFBSs were enriched for H3K4me1/2 in endomesoderm cells (Figure 4A). Overall, 5hmC levels were significantly decreased in cluster 2 regions after differentiation into NPCs [27]. This suggests that a group of 5hmC-enriched enhancers are repressed in mESCs, but selectively become activated during development towards the neuronal or endomesoderm lineage. This implies that other regions in cluster 2 might be activated when ESCs are differentiated into other lineages such as primordial germ cells.

To further determine if the TFBSs in cluster 2 have activating roles in a lineage-specific way, we used chromatin connectivity maps from chromatin interaction analysis with paired-end tagging (ChIA-PET) associated with PolII in ESCs and NPCs [15]. By using the chromatin interaction
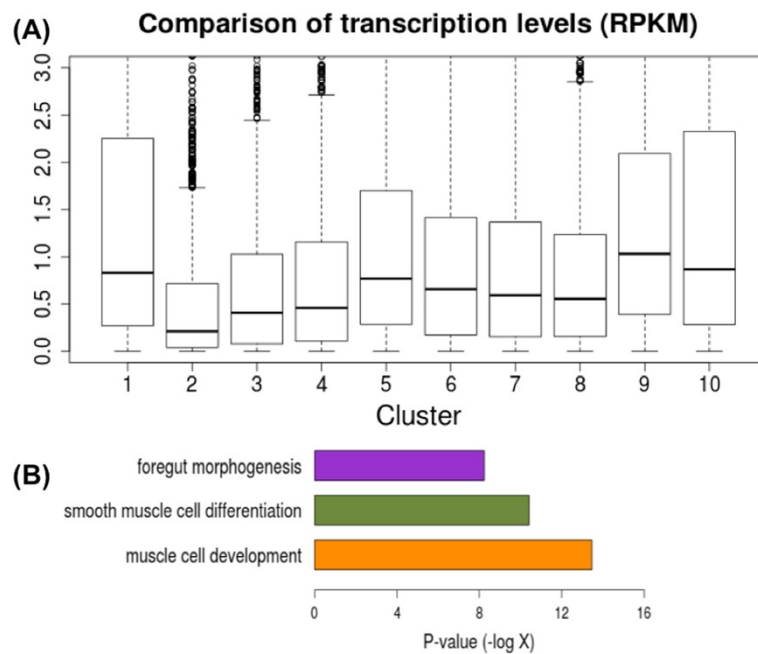
Figure 2 Functional analysis for TFBSs with 5hmC. (A) Comparison of the transcription levels of the nearest genes between cluster 2 and the other 9 clusters of distal TFBSs identified in Figure 1B. The transcription levels (RPKM) were calculated using GROseq data. **(B)** GO analysis for the genes close to TFBSs in cluster 2. Organ development terms are enriched.

information, we mapped the target genes of the TFBSs in cluster 2 that were only selectively activated in NPCs. The target genes in ESCs were slightly downregulated in NPC (as well as in mouse embryo fibroblasts (MEFs)) because only a small portion of them become activated in a lineage-specific manner as shown in Figure 4A. In contrast, the target genes in NPCs become significantly upregulated during the transition from ESCs to NPCs (p-value <0.05). Moreover, the changes were significant compared with the target

genes for clusters (all p-values were <0.02) (Additional file 1: Figure S9). This further supports the notion that TFBSs in cluster 2 become activated in a lineage specific way following embryonic stem cell differentiation.

Cluster 2 regions show enhancer activity in mESCs when devoid of 5hmC

Next, we directly determined if lack of 5hmC activates enhancer activity of the distal TFBSs in mESCs. We
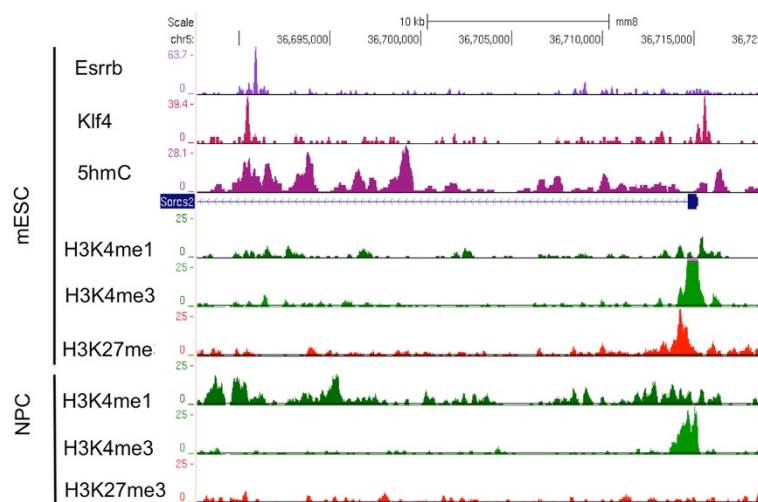

Figure 3 Chromatin organization at the Sorcs2 gene in mESCs. The Klf4 and the Esrrb binding sites in the intron of Sorcs2 gene are enriched for 5hmC. These binding sites are depleted with H3K4me1 in mESCs, but are enriched in NPCs.
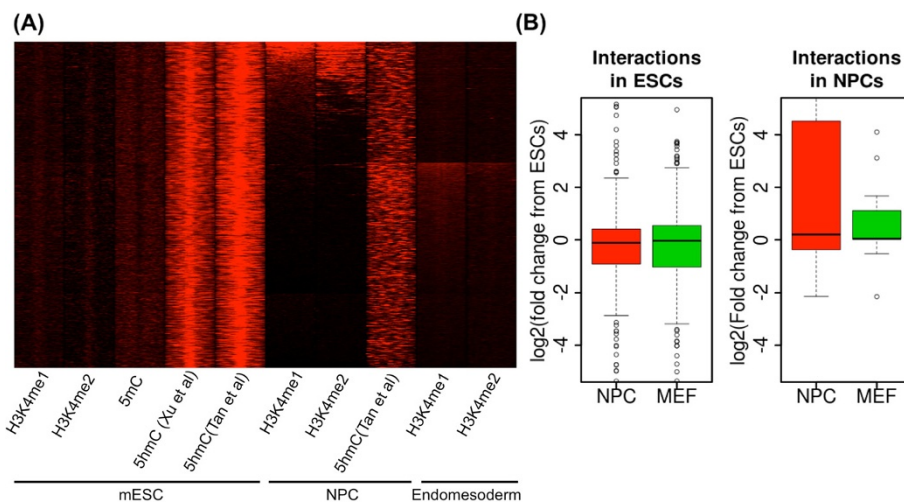
Figure 4 Lineage specific activation of distal TFBSs in cluster 2. (A) The enriched H3K4me1/2 in NPCs or endomesoderm cells suggests the potential lineage specific enhancer activation of the TFBSs in cluster 2 after differentiation. **(B)** The target genes of cluster 2 in NPCs identified using ChIA-PET become significantly upregulated (p-value:0.04) when they gained interactions. The expression change in MEF is compared as a control (p-value:0.12).

selected five highly 5-hydroxymethylated distal TFBSs regions from cluster 2, including the site in *Sorcs2* (TFBS1). This site is also enriched for ChIA-PET [15] reads. We amplified these sequences (±600 bp) by PCR and subcloned them into luciferase reporter plasmids containing a minimal promoter. We found that these DNA sequences indeed possess enhancer activity in mESCs when lacking 5hmC, showing on average 3-fold increased luciferase activity compared to control (Figure 5). This *in vitro* study suggests that 5hmC-enriched distal TFBSs are bona fide enhancers, which are silent in mESCs when marked with 5hmC.

We also investigated if 5hmC at distal regulatory regions has a repressive role using the Tet1 shRNA suppression experiments in mESCs [38]. We found that the target genes of cluster 2 were significantly upregulated (p-value < 0.01) after Tet1 gene suppression, suggestive of repressive roles of 5hmC (Additional file 1: Figure S9). We did not find a similar pattern in Tet2 shRNA-treated mESCs,
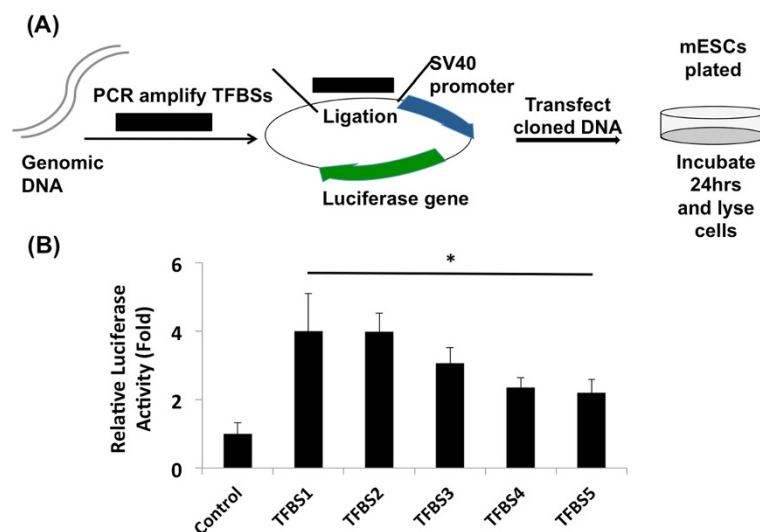

Figure 5 Enhancer activity of distal TFBSs lacking 5hmC in mESCs. (A) Schematic diagram of the experimental setup. **(B)** Luciferase reporter assay for 5hmC-enriched putative enhancer regions (about 600 bp) including distal TFBSs in mESCs. Control: empty vector, TFBS1: 5hmC-enriched Klf4/Esrrb binding site (Figure 3), TFBS2: 5hmC-enriched Esrrb/Tcfcp2l1 binding site, TFBS3: 5hmC-enriched Tcfcp2l1 binding site, TFBS4: 5hmC-enriched E2f1 binding site, TFBS5: 5hmC-enriched Nanog/Sox2. The normalized luciferase activity of control is set as 1. *p-value < 0.05.

possibly due to the fact that Tet2 is dominantly associated with the 5hmC present in gene bodies [38].

Discussion

The field of DNA methylation has expanded recently, with the identification of multiple cytosine variants; 5hmC, 5fC, and 5-carboxylcytosine (5-caC) [28,39]. Among the cytosine variants, 5hmC has been most extensively studies [1,3,9-14]. Although there are genome-wide 5hmC maps in several cell types, our understanding about the functional role of 5hmC remains limited.

The contribution of the 5hmC modification to gene regulation is actively debated. Recent studies found that 5hmC gain is accompanied by H3K27me3 loss at promoters and in the gene body during neurogenesis, suggesting an activating role of 5hmC [40]. On the other hand, the presence of 5hmC at the promoter of bivalently marked genes [8,12,14] and *in vitro* transcription studies revealed a repressive role of 5hmC at promoter regions [41]. However, the role of 5hmC at enhancers has not been investigated thus far.

We observed that enrichment of 5hmC corresponds with the depletion of eRNAs at distal TFBSs. Considering that eRNAs correlate with gene transcription [20,21], we suggest that low levels of 5hmC at enhancers are required for gene expression. Importantly, we found that a subset of distal TFBSs that carry the 5hmC mark in embryonic stem cells become enriched for the activating histone mark (H3K4me1/2) following differentiation into neural progenitors or endomesoderm, suggesting that distal TFBSs with 5hmC are repressed in mESC but become active enhancers in a lineage-specific manner. Indeed, using ChIA-PET interaction information [15], we found that those regions that gained connections to their target genes were significantly upregulated during differentiation compared with the target genes in other clusters. This suggests that their target genes were repressed in ESCs and become selectively activated in a lineage-specific way.

To ascertain if the proposed "silent enhancers" identified above can indeed function as enhancers we employed luciferase reporter assays. We demonstrated that the novel distal elements, characterized by TF binding, high levels of 5hmC, and absence of the H3K4me1 "enhancer" mark, can indeed function as enhancers in mESCs if they are devoid of the 5hmC modification. This experiment is consistent with the notion that 5hmC could inhibit enhancer activity at a subset of distal TFBSs in mESCs.

Our findings are different from the work of Sérandour and colleagues [7], who had suggested an activating role for 5hmC at distal regulatory regions. They identified 5hmC peaks after differentiation which were surrounded by the activating H3K4me2 mark. However, more than 50% the 5hmC peaks they identified were located at genic

regions, where they are known to be associated with gene activation [10,12,38,40,42]. It is also possible that the 5hmC peaks at distal regions are associated with non-coding RNAs such as long non-coding RNAs (lincRNAs) [43]. Sérandour and colleagues also identified 5hmC at distal PPARγ binding sites [33]. Even though Sérandour and colleagues proposed an activating role of 5hmC at these master regulator in adipocytes, only a portion of PPARγ binding sites were enriched for 5hmC [7]. We revisited their data and found that 5hmC was only present at sited lacking PolII occupancy (Additional file 1: Figure S6), indicating that 5hmC at PPARγ binding sites bears repressive roles in mature adipocytes.

In hESCs, we also identified a group of distal DHSs with strong 5hmC but weak H3K4me1 and H3K27ac (Additional file 1: Figure S5). The GROseq levels were significantly weak for the group with 5hmC (Additional file 1: Figure S5). These lines of evidences suggest a general repressive role of 5hmC at distal regulatory regions.

In ESCs, poised enhancers have been suggested to exist at sites where both activating marks (H3K4me1) and repressive marks (H3K27me3) are enriched, but H3K27ac is depleted [23,24]. 5fC is enriched in this type of poised enhancers (H3K4me1[+] and H3K27ac[−]) [30]. In contrast to these poised enhancers, we identify a novel group of enhancers with no activating histone marks (H3K4me1[−] and H3K27ac[−]) but enrichment only for 5hmC. Furthermore, this group is strongly enriched for 5fC, even though cluster 2 lacks the H3K4me1 mark (Figure 1). Our results strongly suggest that 5hmC and 5fC can be epigenetic mark for poised or silent enhancers. As shown in our results, many of these enhancers display activating histone marks only after differentiation has occurred (Figure 4). The existence of 5hmC and 5fC also show the active oxidation dynamics at these sites.

We found that 5hmC was enriched at distal PPARγ binding sites in fully differentiated adipocytes. These findings suggest 5hmC as a new marker for poised enhancers even in absence of H3K4me1 and H3K27me3. Additionally, we also found enriched 5hmC in NPC at the subset of the active TFBSs (except for cluster 2) in mESCs (Additional file 1: Figure S10). This may suggest that active enhancers in mESCs are repressed by 5hmC in NPC to remove the enhancer activities in mESCs.

The majority of cluster 2 regions are CTCF binding sites (Additional file 1: Table S2). In general, 5hmC levels negatively correlated with CTCF occupancy in cluster 2 (Additional file 1: Figure S11). After differentiation into NPCs, 5hmC became depleted at these sites even though the binding CTCF remained. At these sites, we did not observe activating H3K4me1 and H3K4me2 marks. However, it is difficult to discuss the role of 5hmCs at these sites, because CTCF takes part in various regulatory roles including transcriptional activation, repression, as well as

the formation of higher order chromatin structure [44]. The function of 5hmC in mESCs at CTCF binding sites warrants further study.

Conclusions
We report a new repressive role for 5hmC in gene regulatory regions in mESCs. The TFBSs enriched for 5hmCs were depleted for nascent transcripts and activating histone modification marks in human and mouse ESCs. Furthermore, the 5hmC levels were inversely correlated with PolII occupancy in mESCs as well as in fully differentiated adipocytes. Our findings indicate that 5hmC has a repressive role at specific distal regulatory regions and suggest that 5hmC is a new epigenetic mark for silenced enhancers.

Methods
Experimental crocedures
We used genome-wide GROseq maps [18] and ChIP-seq data for chromatin status [17,45], PolII occupancy [17], 5mC [10], and Tet1 occupancy [10] in mESCs for our integrated analysis. We employed H3K4me1/2 data from NPC [17] and endomesoderm cells [37] to analyze the fate of our novel 5hmC regions after differentiation. We also included 5hmC from various independent studies [1,12-14,26,27] for our analysis. Additional file 1: Table S1 summarizes all genome-wide datasets we used in our study.

All ChIP-seq data were normalized to 10 reads per kilobase per million mapped reads (RPKM) [46]. For clustering analysis we used Mev V4.8 [47] and applied the K-means clustering algorithm using the Pearson correlation with absolute distance as a metric. To cluster distal TFBs in mESCs, we used the H3K4me1/2/3, H3K27ac, H3K27me and 5hmC levels and generated applied clustering (K = 10). We showed other epigenetic marks and GROseq and PolII next to the identified clusters.

To study the functional roles of 5hmC in various regulatory regions, we employed binding site data of 13 TFs (Nanog, Oct4, STAT3, Smad1, Sox2, Zfx, c-Myc, n-Myc, Klf4, Esrrb, Tcfcp2l1, E2f1 and CTCF) in mESC [16].

To investigate 5hmC and nascent RNA levels across genes, we divided the genes into promoter (from -1Kbp to 500 bp around the annotated start site), 3′ end (from −500 bp to 500 bp around the annotated termination site), and gene body regions (500 bp from the annotated start site to −500 bp from the annotated termination site). For transcription levels, we calculated RPKM using GROseq reads from 500 bp of the annotated start site to the annotated termination site in order not to include transcriptional pausing at promoters [20,48].

Luciferase reporter assay
Genomic DNA was prepared from R1 mouse embryonic stem cells [49]. About 600 bp genomic fragments for five distal TFBSs in cluster 2 were amplified by PCR with dNTPs and the PCR products ligated into the pGL3-SV40 luciferase vector (Promega). Empty vector (control) or cloned vectors were transfected directly into R1 mESC, together with the pRL-tk vector (Promega) as internal control, using Lipofectamine LTX (Life Technologies). At 24 h after transfection, cells were harvested and lysates subjected to the dual-luciferase reporter assay (Promega). Firefly luciferase activity was measured and normalized to the internal control, Renilla luciferase activity.

Additional file

Additional file 1: Figure S1. 5hmC profile at promoters and enhancers. **Figure S2.** Comparison of the characteristics of each cluster. **Figure S3.** Comparison of the 5hmC patterns for each cluster. **Figure S4.** The 5hmC profile of cluster 2 using TAB-Seq. **Figure S5.** The 5hmC clusters in hESCs. **Figure S6.** The 5hmC clusters in mature adipocytes [10]. **Figure S7.2** The average profiles of TFs at cluster 2. **Figure S8.** The gene expression change for the target genes for each cluster. **Figure S9.** The gene expression changes of the target genes after Tet1 knockdown for each cluster. **Figure S10.** The 5hmC in mESC and NPC at the TFBSs in mESCs. **Figure S11.** 5hmC at CTCF binding sites in cluster 2. **Table S1.** Datasets. **Table S2.** The frequency of transcription factor occupancy in cluster 2.

References
1. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, Helin K: **TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity.** *Nature* 2011, **473**(7347):343–348.
2. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A: **Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.** *Science* 2009, **324**(5929):930–935.
3. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, Min JH, Jin P, Ren B, He C: **Base-resolution analysis of 5-hydroxymethylcytosine in the Mammalian genome.** *Cell* 2012, **149**(6):1368–1380.
4. Kriaucionis S, Heintz N: **The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.** *Science* 2009, **324**(5929):929–930.
5. Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C: **Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine.** *Nat Biotechnol* 2011, **29**(1):68–72.
6. Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N: **MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System.** *Cell* 2012, **151**(7):1417–1430.
7. Serandour AA, Avner S, Oger F, Bizot M, Percevault F, Lucchetti-Miganeh C, Palierne G, Gheeraert C, Barloy-Hubler F, Peron CL, Madigou T, Durand E,

Froguel P, Staels B, Lefebvre P, Métivier R, Eeckhoute J, Salbert G: **Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers.** *Nucleic Acids Res* 2012, **40**(17):8255–8265.

8. Matarese F: **Carrillo-de Santa Pau E, Stunnenberg HG: 5-Hydroxymethylcytosine: a new kid on the epigenetic block?** *Mol Syst Biol* 2011, **7**:562.

9. Szulwach KE, Li X, Li Y, Song CX, Han JW, Kim S, Namburi S, Hermetz K, Kim JJ, Rudd MK, Yoon YS, Ren B, He C, Jin P: **Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells.** *PLoS Genet* 2011, **7**(6):e1002154.

10. Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W: **Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation.** *Nature* 2011, **473**(7347):398–402.

11. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE: **5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells.** *Genome Biol* 2011, **12**(6):R54.

12. Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y: **Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells.** *Genes Dev* 2011, **25**(7):679–684.

13. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, Barbera AJ, Zheng L, Zhang H, Huang S, Min J, Nicholson T, Chen T, Xu G, Shi Y, Zhang K, Shi YG: **Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells.** *Mol Cell* 2011, **42**(4):451–464.

14. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, Tahiliani M, Daley GQ, Liu XS, Ecker JR, Milos PM, Agarwal S, Rao A: **Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells.** *Nature* 2011, **473**(7347):394–397.

15. Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, Lim J, Tai E, Poh HM, Wong E, Mulawadi FH, Sung WK, Nicolis S, Ahituv N, Ruan Y, Wei CL: **Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations.** *Nature* 2013, **504**(7479):306–310.

16. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HH: **Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.** *Cell* 2008, **133**(6):1106–1117.

17. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature* 2007, **448**(7153):553–560.

18. Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, Lis JT: **Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells.** *Genes Dev* 2011, **25**(7):742–754.

19. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES: **Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.** *Nature* 2009, **458**(7235):223–227.

20. Core LJ, Waterfall JJ, Lis JT: **Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.** *Science* 2008, **322**(5909):1845–1848.

21. Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, Glass CK, Rosenfeld MG, Fu XD: **Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA.** *Nature* 2011, **474**(7351):390–394.

22. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B: **Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.** *Nat Genet* 2007, **39**(3):311–318.

23. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: **A unique chromatin signature uncovers early developmental enhancers in humans.** *Nature* 2011, **470**(7333):279–283.

24. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R: **Histone H3K27ac separates active from poised enhancers and predicts developmental state.** *Proc Natl Acad Sci USA* 2010, **107**(50):21931–21936.

25. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR: **Human DNA methylomes at base resolution show widespread epigenomic differences.** *Nature* 2009, **462**(7271):315–322.

26. Wu H, Zhang Y: **Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells.** *Cell Cycle* 2011, **10**(15):2428–2436.

27. Tan L, Xiong L, Xu W, Wu F, Huang N, Xu Y, Kong L, Zheng L, Schwartz L, Shi Y, Shi YG: **Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method.** *Nucleic Acids Res* 2013, **41**(7):e84.

28. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y: **Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.** *Science* 2011, **333**(6047):1300–1303.

29. Maiti A, Drohat AC: **Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites.** *J Biol Chem* 2011, **286**(41):35334–35338.

30. Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin M, Wu H, Gao J, Liu P, Li L, Xu GL, Jin P, He C: **Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming.** *Cell* 2013, **153**(3):678–691.

31. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA: **The NIH Roadmap Epigenomics Mapping Consortium.** *Nat Biotechnol* 2010, **28**(10):1045–1048.

32. Sigova AA, Mullen AC, Molinie B, Gupta S, Orlando DA, Guenther MG, Almada AE, Lin C, Sharp PA, Giallourakis CC, Young RA: **Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells.** *Proc Natl Acad Sci USA* 2013, **110**(8):2876–2881.

33. Nielsen R, Pedersen TA, Hagenbeek D, Moulos P, Siersbaek R, Megens E, Denissov S, Borgesen M, Francoijs KJ, Mandrup S, Stunnenberg HG: **Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis.** *Genes Dev* 2008, **22**(21):2953–2967.

34. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA: **Core transcriptional regulatory circuitry in human embryonic stem cells.** *Cell* 2005, **122**(6):947–956.

35. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G: **GREAT improves functional interpretation of cis-regulatory regions.** *Nat Biotechnol* 2010, **28**(5):495–501.

36. Rezgaoui M, Hermey G, Riedel IB, Hampe W, Schaller HC, Hermans-Borgmeyer I: **Identification of SorCS2, a novel member of the VPS10 domain containing receptor family, prominently expressed in the developing mouse brain.** *Mech Dev* 2001, **100**(2):335–338.

37. Yu P, Xiao S, Xin X, Song CX, Huang W, McDee D, Tanaka T, Wang T, He C, Zhong S: **Spatiotemporal clustering of the epigenome reveals rules of dynamic gene regulation.** *Genome Res* 2013, **23**(2):352–364.

38. Huang Y, Chavez L, Chang X, Wang X, Pastor WA, Kang J, Zepeda-Martinez JA, Pape UJ, Jacobsen SE, Peters B, Rao A: **Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells.** *Proc Natl Acad Sci USA* 2014, **111**(4):1361–1366.

39. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL: **Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA.** *Science* 2011, **333**(6047):1303–1307.

40. Hahn MA, Qiu R, Wu X, Li AX, Zhang H, Wang J, Jui J, Jin SG, Jiang Y, Pfeifer GP, Lu Q: **Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis.** *Cell Rep* 2013, **3**(2):291–300.

41. Robertson J, Robertson AB, Klungland A: **The presence of 5-hydroxymethylcytosine at the gene promoter and not in the gene body negatively regulates gene expression.** *Biochem Biophys Res Commun* 2011, **411**(1):40–43.

42. Kim M, Park YK, Kang TW, Lee SH, Rhee YH, Park JL, Kim HJ, Lee D, Kim SY, Kim YS: **Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage.** *Hum Mol Genet* 2014, **23**(3):657–667.

43. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, Rinn JL, Lander ES, Regev A: **Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the**

conserved multi-exonic structure of lincRNAs. *Nat Biotechnol* 2010, **28**(5):503–510.

44. Nikolaev LG, Akopov SB, Didych DA, Sverdlov ED: **Vertebrate Protein CTCF and its Multiple Roles in a Large-Scale Regulation of Genome Activity.** *Curr Genomics* 2009, **10**(5):294–302.

45. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES: **Genome-scale DNA methylation maps of pluripotent and differentiated cells.** *Nature* 2008, **454**(7205):766–770.

46. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5**(7):621–628.

47. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J: **TM4 microarray software suite.** *Methods Enzymol* 2006, **411**:134–193.

48. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA: **c-Myc regulates transcriptional pause release.** *Cell* 2010, **141**(3):432–445.

49. Nagy A, Rossant J, Nagy R, Abramow-Newerly W, Roder JC: **Derivation of completely cell culture-derived mice from early-passage embryonic stem cells.** *Proc Natl Acad Sci USA* 1993, **90**(18):8424–8428.