

# Nucleosome eviction along with H3K9ac deposition enhances Sox2 binding during human neuroectodermal commitment

Yanhua Du<sup>1,2</sup>, Zhenping Liu<sup>2</sup>, Xinkai Cao<sup>2</sup>, Xiaolong Chen<sup>2</sup>, Zhenyu Chen<sup>3</sup>, Xiaobai Zhang<sup>2</sup>, Xiaoqing Zhang<sup>\*1,3</sup> and Cizhong Jiang<sup>\*1,2</sup>

Neuroectoderm is an important neural precursor. However, chromatin remodeling and its epigenetic regulatory roles during the differentiation of human neuroectodermal cells (hNECs) from human embryonic stem cells (hESCs) remain largely unexplored. Here, we obtained hNECs through directed differentiation from hESCs, and determined chromatin states in the two cell types. Upon differentiation, H2A.Z-mediated nucleosome depletion leads to an open chromatin structure in promoters and upregulates expression of neuroectodermal genes. Increase in H3K9ac signals and decrease in H3K27me3 signals in promoters result in an active chromatin state and activate neuroectodermal genes. Conversely, decrease in H3K9ac signals and increase in H3K27me3 signals in promoters repress pluripotency genes. Moreover, H3K9ac signals facilitate the pluripotency factor Sox2 binding to target sites unique to hNECs. Knockdown of the acetyltransferase Kat2b erases H3K9ac signals, disrupts Sox2 binding, and fails the differentiation. Our results demonstrate a hierarchy of epigenetic regulation of gene expression during the differentiation of hNECs from hESCs through chromatin remodeling.

*Cell Death and Differentiation* (2017) 24, 1121–1131; doi:10.1038/cdd.2017.62; published online 5 May 2017

Chromatin remodeling offers the epigenetic basis for transcriptional regulation and has a pivotal role in many biological processes such as cell differentiation, embryonic development, and so on. Nucleosome is the fundamental repeating structural unit of chromatin.<sup>1</sup> Cell differentiation involves extensive nucleosome reorganization that alters chromatin structure. Lineage-specified cells have a more condensed chromatin structure compared to pluripotency stem cells.<sup>2</sup> Nucleosomes are evicted upstream near transcription start sites (TSSs) or in distant enhancers to activate genes required for lineage commitment during differentiation of stem cells.<sup>3,4,5</sup> Genome-wide comparisons of chromatin structure between mammalian pluripotent stem cells and differentiated cell types reveal that regions of difference in nucleosome occupancy are enriched in the loci associated with pluripotency and differentiation.<sup>6</sup> Thus, dynamics of nucleosome positioning is important to cell fate commitment.

Histone modifications (HMs) are another key impact factor for chromatin structure. HMs exert their epigenetic regulatory roles through changing chromatin state or recruiting non-histone proteins to chromatin.<sup>7</sup> Global changes in HMs during multi-lineages differentiation of human embryonic stem cells (ESCs) have been reported in previous studies.<sup>8,9</sup> Comparisons between human pluripotent stem cells and donor fibroblasts identified lots of regions with differences in H3K4me3 and H3K27me3 signals.<sup>10</sup> Profiling of HMs of mouse ESCs, neural progenitor cells (NPCs), and embryonic fibroblasts revealed

that H3K4me3 and H3K27me3 effectively distinguish genes with different expression levels and reflect lineage potential.<sup>11</sup> A recent HPLC-MS-based quantitative proteomics identified dozens of HM sites in mouse ESCs and NPCs and elaborated the effect of combinational HMs on the differentiation of mouse ESCs to NPCs.<sup>12</sup> The above studies all mainly focused on profiling of HMs and the global changes from ESCs to NPCs. However, the role of HM changes in the differentiation largely remains unclear.

Transcription factors' (TFs) binding to the target sites is closely correlated with HMs for a precise regulation of gene expression. Some TFs contain a protein domain that recognizes specific type of HM. Consequently, the TFs may have an affinity for the specific HM. For example, bromodomain PHD finger TF (BPTF) has a higher affinity to H3K4me3 than H3K4me2 and discriminates against H3K4me1.<sup>13</sup> It was reported that the core pluripotency TF Oct4 binds to the target sites with primed epigenetic signatures during mouse somatic cell reprogramming.<sup>14</sup> Motifs of distinct TFs are enriched in the regions with stage-specific transition during neural differentiation.<sup>15</sup> However, it is unclear which HM(s) have the dominant role and how it regulates TFs' binding. It is also under debate at most time as whether HMs are causal or a consequence of TFs' binding.

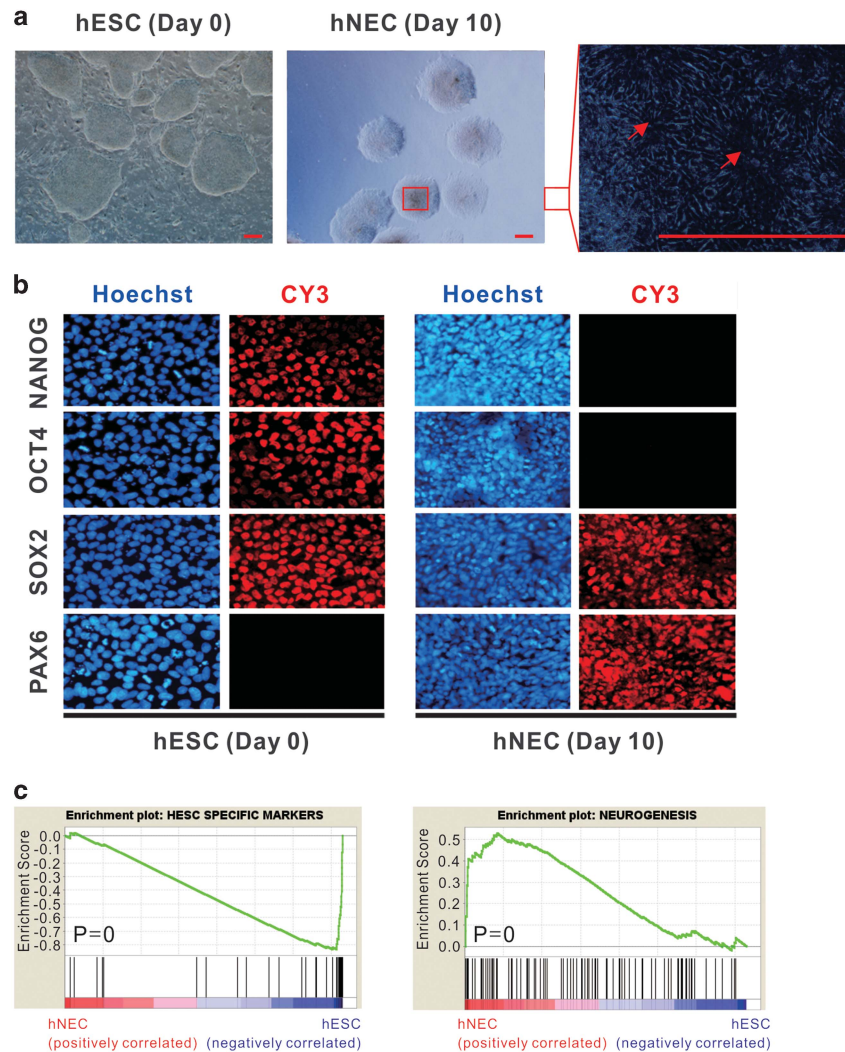
Neuroectoderm specification is of major interest to developmental biology. Neuroectoderm (neuroepithelia) is a very important neural precursor. It develops into neural plate,

<sup>1</sup>Institute of Translational Research, Tongji Hospital, Tongji University, Shanghai 200065, China; <sup>2</sup>The School of Life Sciences and Technology, Shanghai Key Laboratory of Signaling and Disease Research, The Collaborative Innovation Center for Brain Science, Tongji University, Shanghai 200092, China and <sup>3</sup>Tongji University School of Medicine, Shanghai 200092, China

\*Corresponding author: X Zhang, Tongji University School of Medicine, 1239 Siping Road, Room 702, Shanghai 200092, China. Tel: +86-21-6598-5003; Fax: +86-21-6598-5003; E-mail: xqzhang@tongji.edu.cn

or C Jiang, School of Life Sciences and Technology, Tongji University, 1239 Siping Road, Shanghai 200092, China. Tel: +86-21-6598-1193; Fax: +86-21-6598-1041; E-mail: czjiang@tongji.edu.cn

Received 21.2.17; revised 28.3.17; accepted 31.3.17; Edited by N Bazan; published online 05.5.2017



**Figure 1** Expression profiles along hESCs differentiating to hNECs. (a) Microscopy images showing cell morphology changes between hESCs and hNECs. The right panel is an enlargement of hNEC colony marked by a red box. Arrows indicating that the differentiated hNECs are organized in rosettes. Scale bars, 100  $\mu$ m. (b) Immunofluorescent staining showing expression of TFs Nanog, Oct4, Sox2 and Pax6 in hESCs and hNECs, respectively. (c) GSEA results: hESC-specific genes are downregulated in hNECs whereas neurogenesis genes are upregulated

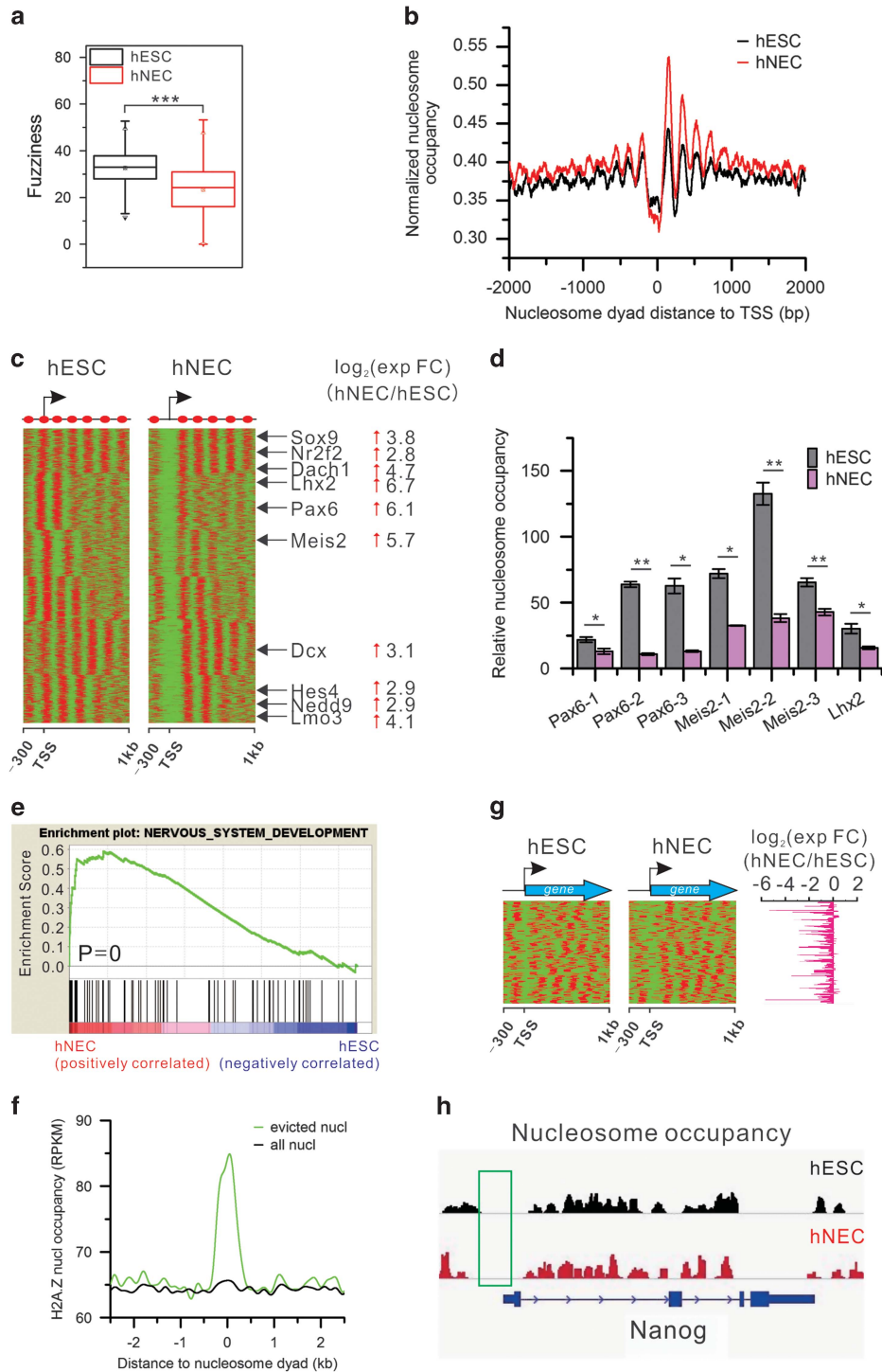
providing the source of the central nervous system.<sup>16</sup> Moreover, neuroectoderm has potentials to differentiate into neurons, astrocytes, and oligodendrocytes. Therefore, the differentiation of ESCs to neuroectoderm is of great significance for understanding of neural disease, for example, primitive neuroectodermal tumors. Epigenetic regulation has a critical role in the differentiation. For example, a subset of poised enhancers gained active chromatin signatures (H3K27ac) during differentiation of hESCs to neuroepithelium.<sup>17</sup> However, to date, the epigenetic regulatory mechanisms underlying neuroectoderm specification remains elusive.

To unveil the patterns and roles of chromatin remodeling during the differentiation of human ESCs (hESCs) to human neuroectodermal cells (hNECs), we employed the differentiation method of hESCs to a synchronized population of hNECs.<sup>18,19</sup> This differentiation process resembles *in vivo* neural development and serves a model to study the mechanisms underlying human neuroectoderm specification.<sup>20</sup> For example, Pax6 was identified as a transcriptional determinant

of human neuroectoderm through this neural differentiation tool.<sup>21</sup> In this study, we profiled transcriptome, nucleosome occupancy, core histone modifications, and genomic bindings of TF Sox2 for hESCs and hNECs. The integrative analyses of these data show that extensive nucleosome repositioning and histone modification changes occur upon differentiation. H2A.Z-mediated nucleosome disassembly in promoters leads to an open chromatin structure to activate transcription of neuroectodermal genes. Moreover, H3K9ac signals deposited by acetyltransferase Kat2b are essential to the differentiation through enhancing recruitment neuroectodermal factors Sox2 and Pax6 to their target sites.

## Results

**Expression profiles upon differentiation.** We employed a chemically defined protocol to make directed differentiation of hESCs to nearly pure hNECs.<sup>18,19</sup> The synchronized



**Figure 2** Nucleosome remodeling upon the differentiation. (a) Nucleosome fuzziness is significantly decreased in hNECs. ( $P$ -value =  $2.2e - 16$ , Student's  $t$ -test). (b) Composite nucleosome distribution around TSS. (c) NDR formation in hNECs around TSS of a set of genes. Representative neuroectodermal genes are listed in the middle whose expression change folds are indicated at right. Arrows indicate upregulation in hNECs. (d) ChIP-qPCR confirmation of decreased nucleosome occupancy in NDRs of the selected neuroectodermal genes. Error bars represent SEM. (\* $P < 0.05$ , \*\* $P < 0.01$ , Student's  $t$ -test). (e) GSEA showing that nervous system development genes are enriched in the genes in c with an NDR in their promoters and upregulated in hNECs. (f) Nucleosomes evicted in hNECs are abundant with H2A.Z nucleosomes. (g) NDRs remain around TSS of pluripotency genes whose expression levels are greatly decreased in hNECs. (h) Track view of NDRs around TSS of pluripotency TF Nanog maintained in both hESCs and hNECs. Green box indicates the NDRs

population of hNECs have a characteristic of rosette formation that is morphologically distinct from hESCs (Figure 1a). The qPCR results show that the mRNA levels of core pluripotency genes Oct4 (also known as Pou5f1) and Nanog are significantly decreased in hNECs whereas the hNEC determinant factor Pax6 is opposite (Supplementary Figure S1a). Immunofluorescence assay confirmed the similar results at protein levels (Figure 1b). Of note, Sox2 has a retention of high protein levels in both hESCs and hNECs. This is consistent with the critical roles of Sox2 in both ESCs and NPCs.<sup>22,23</sup> We'll discuss more about the functions of Sox2 on the differentiation of hNECs hereinafter.

Gene set enrichment analysis (GSEA)<sup>24</sup> revealed that hESC-specific marker genes are significantly repressed during the differentiation, whereas the genes associated with neurogenesis and central nervous system development are significantly upregulated (Figure 1c and Supplementary Figure S1b). The profiling of changes in global gene expression identified 2477 significantly differentially expressed genes (DEGs), 1033 upregulated and 1444 down-regulated in hNECs (Supplementary Figure 1c). Gene ontology (GO) analysis found that the upregulated DEGs are enriched for terms related to cell differentiation and neurogenesis (Supplementary Figure 1d). Taken together, the expression profiles consist with the corresponding cell identities. The results confirmed the purity of differentiated hNECs at the molecular levels.

**H2A.Z-mediated nucleosome eviction.** We scanned the genome with a 10-kb window and calculated nucleosome occupancy for each window. Nucleosome occupancy of the biological replicates is highly reproducible (Supplementary Figure S2a). There occur global extensive changes in nucleosome occupancy in the differentiation with significant occupancy increase in most genomic regions in hNECs (Supplementary Figure S2b). We obtained the same results with different offsets and window sizes (data not shown here). Moreover, the nucleosome positioning becomes more fixed in hNECs than in hESCs (Figure 2a). As a result, chromatin structure becomes more condensed and stable in hNECs. This is consistent with the chromatin structures in pluripotent stem cells and differentiated cells, respectively.<sup>2</sup>

We next predicted nucleosome positions using nucleosome prediction tool GeneTrack.<sup>25</sup> Comparisons found that only 1% of nucleosomes are fixed without changing their positions upon differentiation, 6% are disassembled or reassembled, and the rest shift. Further analysis show that nucleosome loss is enriched in promoters (Supplementary Figure S2c). This finding suggests that there are extensive nucleosome repositioning during the differentiation.

Nucleosome organization in promoters has a critical role in regulation of gene expression.<sup>26</sup> There exists a canonical nucleosome organization of -1, nucleosome depletion region (NDR), +1, +2, *et al.* nucleosomes around TSS in both cell types. The average nucleosome occupancy in promoters is lower in hESCs than hNECs. Surprisingly, there is a small peak in the NDR in hESCs indicating that NDRs of some genes are occupied by nucleosomes in hESCs but free of nucleosomes in hNECs (Figure 2b). The nucleosome occupancy in the NDRs (-150-50 bp of TSS) is negatively

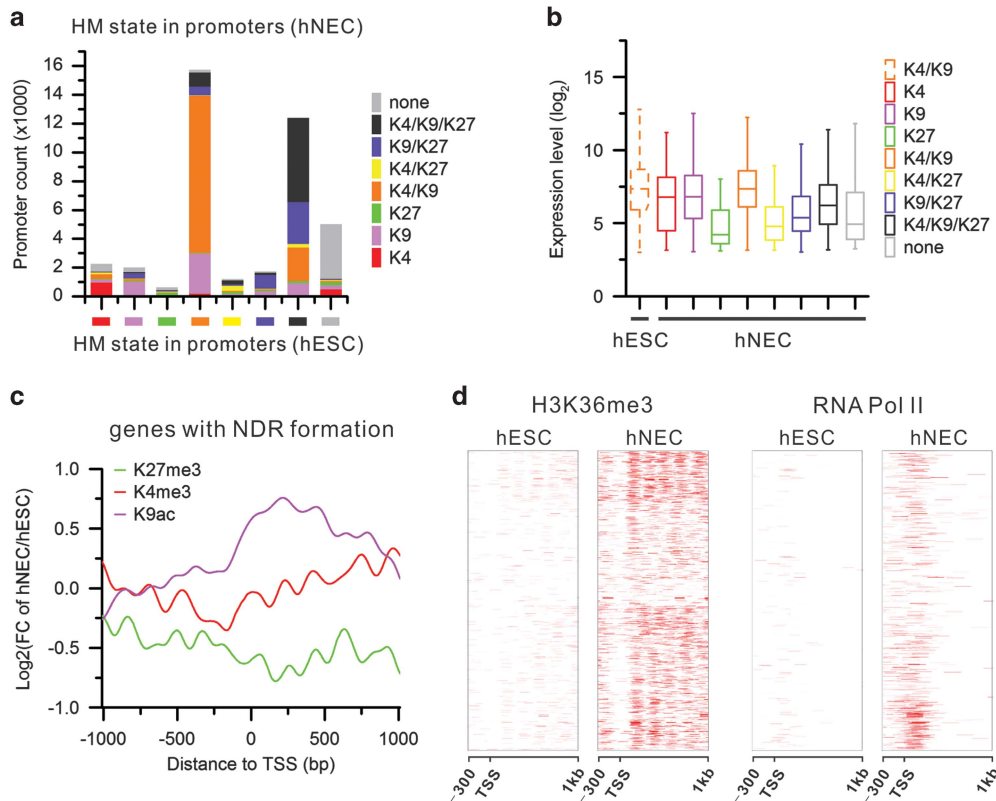
correlated with gene expression levels (Supplementary Figure S2d). This suggests the key roles of NDRs in regulating differentiation.

To address the role of NDR formation in the differentiation of hNECs, we identified genes containing an NDR in hNECs that is occupied by a nucleosome in hESCs. Nucleosome arrangement around TSS of these genes is similar in both cell types except for an occupied NDR in hESCs and an unoccupied NDR in hNECs (Figure 2c). Intriguingly, this group of genes include many neuroectodermal genes such as Lhx2, Pax6, Meis2 and so on. As a result, their expression levels are significantly increased. ChIP-qPCR results confirmed the significantly reduced nucleosome occupancy in the NDRs (Figure 2d). GSEA results further show that 'nervous system development' gene set is activated upon differentiation (Figure 2e). Remarkably, H2A.Z nucleosomes are enriched on the evicted nucleosomes that form the NDRs in hNECs (Figure 2f). This suggests that H2A.Z nucleosomes mediate NDR formation in promoters of neuroectodermal genes to facilitate the differentiation of hNECs. As contrast, we collected pluripotency genes and examined nucleosome organization around TSS. Unexpectedly, NDRs retain unoccupied during the differentiation (Figures 2g and h; Supplementary Figure S2e). However, their expression levels are significantly decreased (Figure 2g). This indicates that other factors other than nucleosome occupancy in NDRs regulate their activity.

We further analyzed the distribution of evicted nucleosomes and H2A.Z nucleosomes around TSS. The results show that evicted nucleosomes are enriched in the canonical -1, NDR, +1, +2 nucleosomes. Consistently, H2A.Z nucleosomes are also enriched in this region (Supplementary Figure S2f). In addition, only 1.2% of evicted nucleosomes and 10.4% of H2A.Z nucleosomes are located around TSS, respectively. These findings suggest that nucleosome eviction is not restricted to the promoter regions on the genome scale.

**Establishment of a permissive chromatin state in promoters.** To uncover how HMs change and regulate gene expression during the differentiation, we generated genome-wide occupancy maps of core HMs using high-throughput ChIP-seq technology with high reproducibility (Supplementary Figure S3a). We next categorized promoters by HM marks. The majority of promoters are marked by H3K4me3/H3K9ac or H3K4me3/H3K9ac/H3K27me3 in both cell types. The prominent difference is that more promoters are marked by H3K9ac only or H3K9ac/H3K27me3 in hNECs than hESCs whereas promoters marked by H3K4me3/H3K9ac/H3K27me3 are opposite (Supplementary Figure S3b). This implies that H3K9ac likely has a critical role in the differentiation of hNECs. Analysis of HM dynamics shows that the majority of promoters remain their original chromatin state in the differentiation of hNECs from hESCs (Figure 3a and Supplementary Figure S3c). Moreover, the increased promoters marked by H3K9ac only in hNECs are from promoters marked by H4K3me3/H3K9ac in hESCs. The increased promoters marked by H3K9ac/H3K27me3 in hNECs are from promoters marked by H4K3me3/H3K9ac/H3K27me3 in hESCs. As a result, gene expression levels increase as active HM signals





**Figure 3** Establishment of a permissive chromatin state in the promoter regions. (a) Statistics summary of each category of promoters defined by histone modification mark (s) in hESCs changing into other categories in hNECs. (b) Changes in expression levels of genes with H3K4me3/H3K9ac in hESCs upon differentiation to hNECs. Gain of repressive H3K27me3 is associated with decreased expression. (c) Fold change of core histone modification signals in the promoters with NDR formation upon differentiation. (d) Increased H3K36me3 signals and RNA Pol II binding in the genes with NDR formation in the promoters upon differentiation

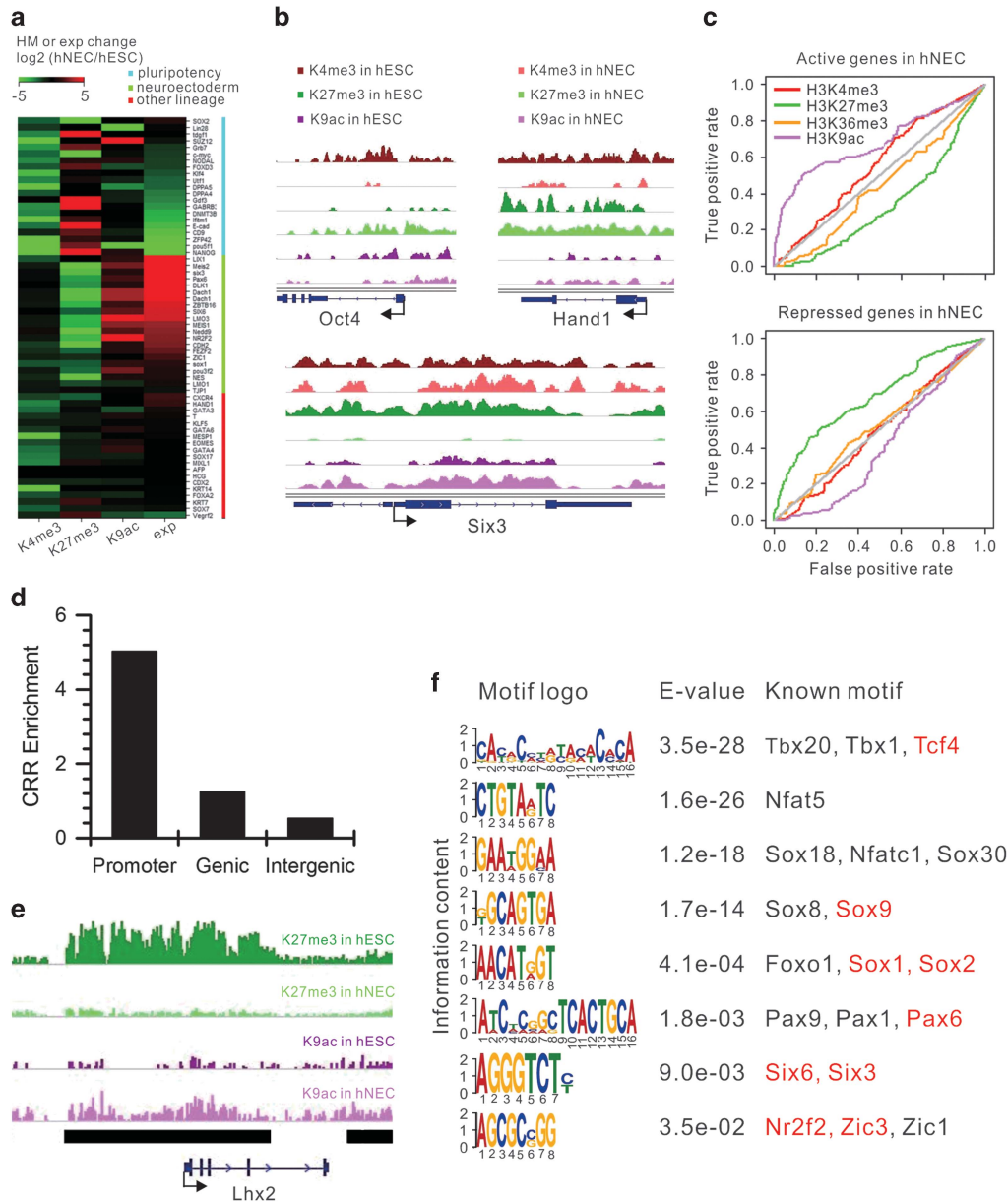
increase and/or repressive HM signals decrease, vice versa (Figure 3b and Supplementary Figure S3d).

We next specifically examined HM dynamics in the promoters with NDR formation during the differentiation. The active HM signals largely increase around TSS from hESCs to hNECs, especially H3K9ac levels are increased to approximately twofolds. Conversely, the repressive H3K27me3 is decreased (Figure 3c). There is also enrichment of RNA Pol II around the TSS and of H3K36me3 across the gene body that indicates active transcription of these genes (Figure 3d). This suggests that nucleosome eviction together with HM dynamics set up an open chromatin structure in the promoter regions in neuroectodermal genes for transcription to facilitate the differentiation of hNECs from hESCs.

**H3K9ac and H3K27me3 dynamics accurately regulates gene activity of distinct lineages.** To address which HM(s) have a dominant role in control of gene activity in the differentiation of hNECs, we collected three sets of genes: pluripotency genes for hESCs, neuroectodermal genes for hNECs, and other lineage-specific genes for comparison, and examines HM dynamics in promoters and gene expression change. The results show that H3K4me3 levels largely decrease in pluripotency genes whereas H3K27me3 levels increase. H3K9ac levels change very little. Concomitantly, the gene expressions are downregulated (Figures 4a and b). This suggests that H3K27me3 and H3K4me3 have a pivotal role

in regulation of pluripotency genes. Contrarily, H3K27me3 signals greatly decrease in neuroectodermal genes, whereas H3K9ac signals increase. H3K4me3 signals remain unchanged. Consequently, the gene expressions are upregulated (Figures 4a and b). Moreover, H3K9ac and H3K27me3 levels are the best predictive for gene activity in hNECs among the core HMs (Figure 4c). This indicates that H3K27me3 and H3K9ac have a dominant role in regulation of neuroectodermal genes. In contrast, there is very little change in both H3K27me3 and H3K9ac levels in other lineage-specific genes in concordance with unchanged gene expression levels (Figures 4a and b). This implies that other lineage-specific genes remain unchanged expression levels under the context of unchanged chromatin structure in promoters. ChIP-qPCR results confirmed the aforementioned distinct HM dynamics in promoters of the representative genes (Supplementary Figure S4a).

We further quantified change in H3K9ac and H3K27me3 signals in promoters upon differentiation by calculating the change in levels of the two HMs:  $\log_2(\text{fold change of H3K9ac}_{\text{hNECs/hESCs}}) - \log_2(\text{fold change of H3K27me3}_{\text{hNECs/hESCs}})$  that is defined as Histone Modification Index (HMI). The distribution of HMI in genes is a normal distribution (Supplementary Figure S4b). The net effect of  $\text{HMI} \leq -1$  is that increase in H3K9ac in promoters of these genes is equal to or less than two folds of increase in H3K27me3. Consistently, the expression levels of genes with  $\text{HMI} \leq -1$



**Figure 4** H3K9ac and H3K27me3 dynamics accurately regulates gene activity. (a) Changes of core histone modification signals in the promoters of three categories of genes ordered by expression changes within each category. (b) Track view of core histone modification signals in the representative genes of each category (Oct4: pluripotency, Hand1: other lineage, Six3: neuroectoderm). (c) Receiver operating characteristic (ROC) curves of four histone modifications as predictors of gene activity in hNECs. (d) Enrichment of chromatin remodeling regions (CRRs) in different genomic regions. The enrichment of CRRs in promoters is equal to the number of CRRs in promoters normalized by the length of promoters. The enrichment of CRRs in gene body and intergenic regions are calculated in the same way. (e) Track view of a representative CRR (labeled by the black line) in the promoter of neuroectodermal TF Lhx2. (f) Sequence motifs found in CRRs through *de novo* enrichment analysis. Corresponding E-values are indicated for each motif. The last column lists the TFs associated with similar motifs. Neuroectodermal TFs are in red

significantly decrease in hNECs. Conversely, the net effect of  $HMI \geq 1$  is that increase in H3K9ac in promoters of these genes is equal to or more than two folds of increase in H3K27me3. The expression levels of genes with  $HMI \geq 1$  significantly increase in hNECs. In contrast, there is no significant change in the expression levels of the rest of genes (Supplementary Figure S4c). This suggests that the combinational H3K9ac and H3K27me3 marks serve an appropriate predictive of global gene activity and have a critical role in the differentiation of hNECs from hESCs.

As increase in H3K9ac and decrease in H3K27me3 in promoters have a dominant role in activation of neuroectodermal genes (Figure 4a), we extended the similar analysis to the whole genome by identifying all regions prominent increase in H3K9ac or decrease in H3K27me3 upon differentiation, which are defined as chromatin remodeling regions (CRRs, see METHODS for details). There are total 7389 CRRs most of which have a length of 2000–6000 bp (Supplementary Figure S4d). Interestingly, CRRs are enriched in promoters (Figures 4d and e; Supplementary Figure S4e).

Moreover, CRRs contain binding motifs of many TFs associated with neural differentiation and development, including Pax6, Six6, and Sox family (Figure 4f). Collectively, H3K9ac and H3K27me3 execute their epigenetic regulatory functions during the differentiation through chromatin remodeling mainly in promoters and consequently likely controlling accessibility of binding motifs of key TFs.

**Sox2 binds to hNEC-specific sites upon the differentiation.** TF Sox2 has important roles in both ESCs and progenitors of multiple lineages.<sup>22,23</sup> How Sox2 executes its distinct functions, especially in the differentiation of hNECs, remains largely unknown. Our highly reproducible Sox2 ChIP-seq data (Supplementary Figure S5a) identified 12,623 binding sites in hESCs and 4875 sites in hNECs. Most are unique in each cell type except 411 sites common in both cell types (Figure 5a). The binding site sequences are abundant with Sox2-binding motif that resembles the canonical Sox2 motif in the Cistrome database<sup>27</sup> (Supplementary Figure S5b). This suggests that our Sox2 ChIP-seq data capture its bona fide target sites. *De novo* motif discovery further identified motifs of other pluripotency TFs, including Oct4, Nanog, and so on, in hESC-specific Sox2-binding sites. In contrast, motifs of other neural TFs, including Nkx2.1, Sox9, Pax6, and so on, are enriched in hNEC-specific Sox2-binding sites (Figure 5b). This together indicates that Sox2 exerts its distinct functions in pluripotency maintenance and multilineage progenitor commitment by binding to distinct target sites with involvement of other corresponding co-factors.

**Kat2b deposits H3K9ac to recruit Sox2 to its target sites.** To understand how chromatin remodeling regulates Sox2 binding to its target sites upon differentiation, we examined nucleosome occupancy around hNEC-specific Sox2-binding sites and found no change during the differentiation (Supplementary Figure S5c). However, H3K9ac levels are largely increased on the sites whereas H3K27me3 levels are decreased (Figure 5c). Moreover, Sox2-binding signals are enriched in CRRs (Figure 5d). This implies that H3K9ac and H3K27me3 are a dominant factor for recruitment of Sox2 to the target sites other than nucleosome occupancy.

Acetyltransferase Kat2b is the key enzyme responsible for H3K9ac deposition. Coincidentally, Kat2b is significantly upregulated upon differentiation. Thus, we hypothesized that Kat2b may be the key factor to regulate Sox2 binding by depositing H3K9ac upon differentiation. To test this, we knocked down Kat2b by shRNAs (Supplementary Figure S5d). As knockdown of Kat2b, H3K9ac levels are significantly decreased at Sox2-binding sites, and Sox2 occupancy is also significantly reduced. Similarly, Kat2b knockdown also significantly reduces H3K9ac levels at Pax6 target sites and consequently Pax6 binding signals are precipitously decreased (Figure 5e). As a result, Kat2b knockdown fails the differentiation of hNECs from hESCs (Supplementary Figure S5e).

## Discussion

Chromatin remodeling has a critical role in cell differentiation. In this study, the results show there occur extensive nucleosome repositioning and HM dynamics during the

differentiation of hNECs from hESCs. H2A.Z mediates nucleosome eviction to form NDRs in the promoters of neuroectodermal genes. Acetyltransferase Kat2b deposits H3K9ac in the genes in concomitance with decrease in repressive H3K27me3 levels. The resultant open chromatin structure in the promoters activates neuroectodermal gene transcription. Meanwhile, the increased H3K9ac signals help recruit Sox2 to its target sites unique to hNECs. As a result, hESCs are directly differentiated to hNECs (Figure 6). These findings reveal the dynamics of chromatin structure and its epigenetic regulatory roles in the differentiation of hNECs from hESCs. This indicates that the fine-tuning of chromatin structure is critical to lineage-commitment gene regulation.

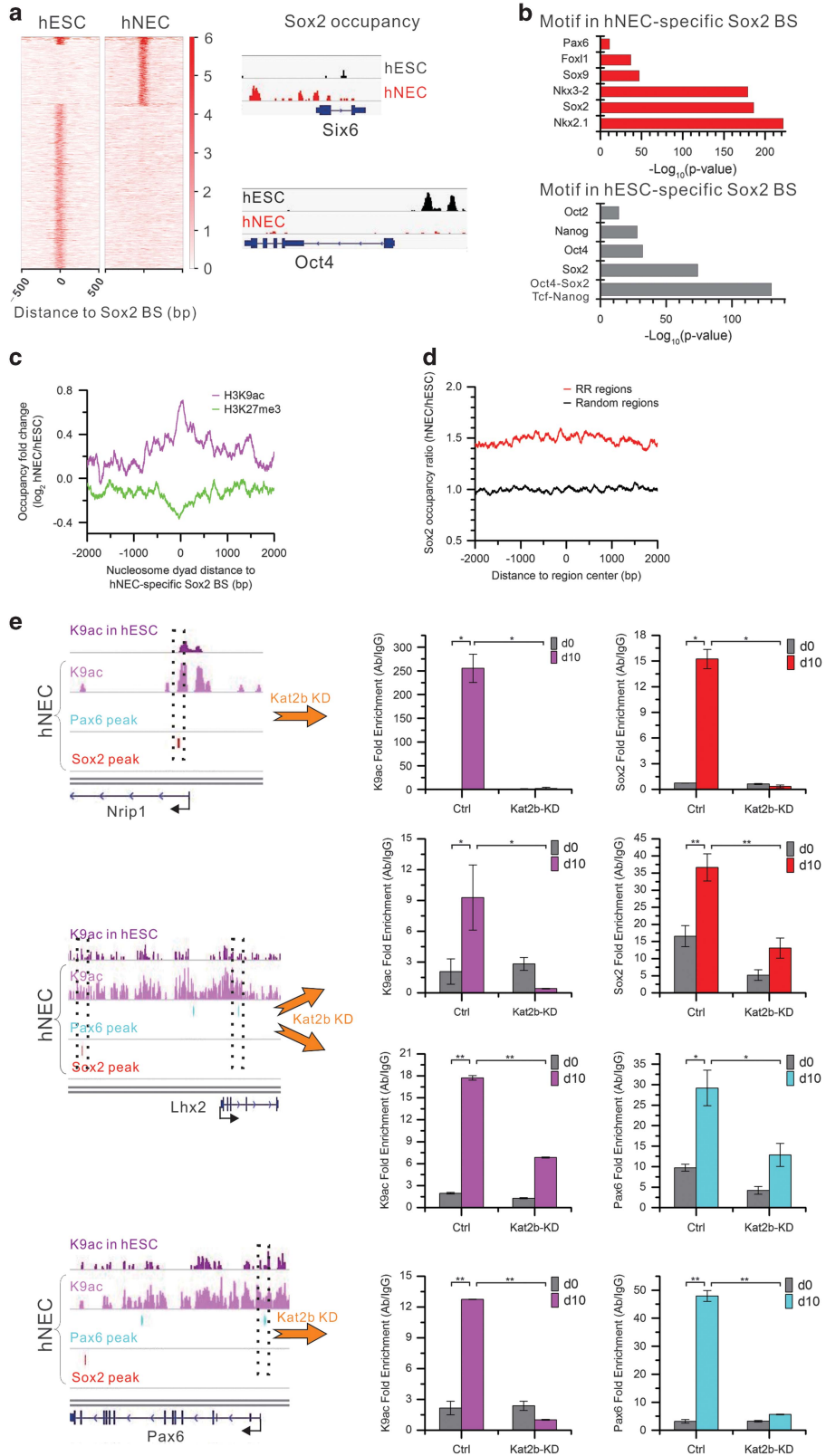
Dynamic changes in chromatin modifications have crucial roles in cell differentiation. A previous study illustrated that H3K4me3 at promoters remains largely invariant during hESCs differentiating to a mesendodermal lineage. In contrast, switch between H3K27ac and H3K27me3 at promoters is critical to the differentiation by repressing pluripotency genes and activating mesendodermal genes.<sup>28</sup> Intriguingly, our findings suggest that dynamic changes in H3K9ac and H3K27me3 at promoters are important to the differentiation of hNECs from hESCs. Consistently, H3K9ac levels at promoters are critical to glial and neuronal commitment in *Drosophila* embryonic neural development.<sup>4,5,29</sup> Collectively, marked changes in distinct combinational histone modifications in cell fate commitment tend to be cell-type specific.

Nucleosome occupancy is fundamental for chromatin structure. Nucleosome eviction leads to an open chromatin structure increasing accessibility of DNA sequences and therefore has an important role in the differentiation process. It was reported that chromatin remodeling complexes SWI/SNF and INO80 participate nucleosome depletion in promoters during differentiation toward the endoderm/hepatic fate.<sup>3</sup> Similarly, we observed H2A.Z-mediated nucleosome eviction in promoters upon differentiation of hNECs than significantly increase transcription levels of neuroectodermal genes. However, the detailed mechanism of which chromatin remodeling complexes orchestrate to regulate nucleosome eviction requires future studies.

Interplay between TFs and histone post-translational modifications precisely regulates transcription. Chromatin modifications enhance recruitment of TFs to certain genomic regions. For example, TFs and co-activators favor binding to transcriptional enhancers defined by H3K4me1 enrichment. Extended H3K27ac domains are abundant of motifs for master TFs in the respective cell types.<sup>30,31</sup> Here, we reveal that H3K9ac marks at the hNEC-specific binding sites of neuroectodermal factors Sox2 and Pax6 are essential for their binding. Knockdown of acetyltransferase Kat2b largely reduces H3K9ac levels, disrupts bindings of Sox2 and Pax6, and consequently fails neuroectodermal differentiation. This greatly improves our understanding of regulatory roles of chromatin remodeling in the differentiation of hNECs from hESCs.

## Materials and Methods

**Differentiation of hNECs from hESCs.** Human ESCs (H9, passages 18–21) were differentiated into hNECs as described previously.<sup>19,21</sup> Briefly, hESC colonies were cultured on irradiated mouse embryonic fibroblasts (MEFs) in ESC growth medium (Sterilely combine 392.5 ml DMEM/F12, 100 ml Knockout serum replacer, 5 ml MEM nonessential amino acids solution, 2.5 ml of 200 mM L-glutamine





solution (final concentration of 1 mM), and 3.5 ml 14.3 M  $\beta$ -Mercaptoethanol (final concentration of 0.1 mM) supplemented with fresh 4 ng/ml recombinant human FGF basic (bFGF, Invitrogen, Waltham, MA, USA, Cat. No. 13256-029). Then, the colonies were detached from MEFs and formed aggregates to initiate the neural differentiation procedure by withdrawing bFGF. After 4 days of suspension culture, the ESC growth medium was replaced with neural induction medium (Sterilely combine 489 ml of DMEM/F12, 5 ml N2 supplement, 5 ml MEM nonessential amino acids solution, and 1 ml of 1 mg/ml Heparin) to guide neuroectodermal specification. Then transferred the suspension cells to the 6-well plates without feeder on the sixth day. After another 4 days of adherent growth, we got hNECs on day 10.

**Immunofluorescence staining.** Cells were fixed by 4% paraformaldehyde at room temperature for 10 min, followed by 0.1% Triton X-100 permeabilization for 10 min and 1% BSA blocking for 1 h. They were then incubated with primary antibodies against Oct4 (1:500, Santa Cruz, Dallas, TX, USA), Nanog (1:500, R&D, Minneapolis, MN, USA), Sox2 (1:500, R&D) or Pax6 (1:500, DSHB, Iowa City, IA, USA) at 4 °C overnight. The next day, cells were incubated with secondary antibodies against mouse or goat IgG (Life Technologies, Waltham, MA, USA) accordingly. Nuclei were counterstained with Hoechst (Invitrogen), and confocal images were taken on the Zeiss LSM 710 microscope (Goettingen, Germany).

**mRNA extraction and qRT-PCR.** Total RNA was isolated using the Trizol reagent (Invitrogen) and RNA concentration was determined by NanoDrop 2000 c (Thermo Scientific, Waltham, MA, USA). Overall, 1  $\mu$ g of total RNA was reversely transcribed into cDNA using SuperScript III (Invitrogen) and subjected to real-time PCR (Bio-Rad, Hercules, CA, USA, CFX Connect Real-Time System) using the Ssofast EvaGreen kit (Bio-Rad). The real-time PCR primers were listed in the Supplementary Table S1.

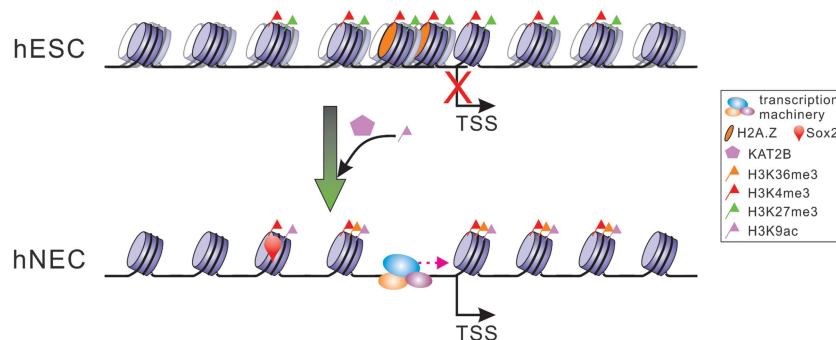
**MNase-seq and ChIP-seq for histone modifications.** Overall,  $\sim 2 \times 10^7$  cells were crosslinked with 1% formaldehyde for 10 min at room temperature, and crosslinking was terminated by addition of 125 mM glycine. Cells were harvested for lysis to isolate nuclei. Suspend nuclei in 500  $\mu$ l of MNase digestion buffer (10 mM Tris-HCl (pH 7.5), 15 mM NaCl, 60 mM KCl, 1 mM CaCl<sub>2</sub>, 0.15 mM spermine, 0.5 mM spermidine, 1  $\times$  EDTA free protease inhibitor cocktail). Then digest nuclei with 2.5  $\mu$ l of micrococcal nuclease (NEB# M0247S, 2000 gel units/ $\mu$ l) at 37 °C for 20 min. Then put the samples on ice and add EDTA to a final concentration of 10 mM to stop MNase digestion. Crosslinking was then reversed for at least 6 h at 65 °C along with proteinase K digestion. Nucleosomal DNA was extracted using phenol-chloroform and purified on 1.5% agarose gels.

A total of 10–20  $\mu$ g of nucleosomal DNA was used for chromatin immunoprecipitation with 3–5  $\mu$ g of HM antibodies as described in 400- $\mu$ l ChIP buffer (2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl (pH 8.1), 0.1% Triton X-100, 1 $\times$ EDTA free protease inhibitor cocktail). Antibodies used are from Abcam (Cambridge, MA, USA): H3K4me3 (ab8580), H3K27me3 (ab6002), H3K36me3 (ab9050), H3K9ac (ab10812). The mixture was incubated overnight at 4 °C with rotation. Then, 30  $\mu$ l of ChIP-Grade Protein G Magnetic Beads (CST; #9006) were added in the mixture for another 2 h. After that, beads were sequentially washed three times with low salt buffer (2 mM EDTA, 20 mM Tris-HCl (pH 8.1), 0.1% SDS, 1% Triton X-100, 150 mM NaCl), once with high salt buffer (2 mM EDTA, 20 mM Tris-HCl (pH 8.1), 0.1% SDS, 1% Triton X-100, 500 mM NaCl), once with LiCl wash buffer (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl (pH 8.1)), and once with TE buffer (10 mM Tris-HCl (pH 8.1), 1 mM EDTA). Then beads were suspended in elution buffer (50 mM Tris-HCl (pH 8.1), 10 mM EDTA, 0.1–0.5% SDS) at 65 °C for 1 h.

The purified mononucleosomal DNA by MNase digestion and ChIP'ed by HM antibodies were subjected to massively parallel DNA sequencing on Illumina HiSeq2000 platform (San Diego, CA, USA) using 49 bp single end protocol.

**ChIP-seq for Sox2 and RNA Pol II.** Genomic DNA was crosslinked and extracted as described above MNase-seq. The purified genomic DNA was suspended in 500  $\mu$ l of sonication buffer (1 $\times$ XPBS, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 1  $\times$  EDTA free protease inhibitor cocktail), and sonicated with 6 rounds of 30-s on and 30-s off on ice using XL-2000 Misonix sonicator with power output of 7 Watts. It is critical that the average length of the sheared chromatin is about 250 bp, with length ranging from 150–500 bp. The fragmented DNA was immunoprecipitated with 3–5  $\mu$ g of specified antibody (anti-Sox2: Abcam ab59776, anti-Pol II: Abcam ab817) as described in the above ChIP-seq for HMs. The ChIP'ed DNA was subjected to massively parallel DNA sequencing on Illumina HiSeq2000 platform (San Diego, CA, USA) using 49 bp single end protocol.

**ChIP-qPCR.** ChIP-qPCR was conducted according to user manual of ChampionChIP PCR Array from SABiosciences (Germantown, MD, USA). Briefly, DNA fragments were prepared as above MNase-seq or ChIP-seq. Relative signal abundance in regions of interest in sample DNA was measured by qPCR using Power SYBR. Fragmented genomic DNA or IgG-immunoprecipitated DNA was used as control sample. Relative signal enrichment was calculated using  $\Delta\Delta$ Ct method by normalizing Ct values against control sample. The ChIP-qPCR primers were listed in the Supplementary Table S2.



**Figure 6** Model of how Kat2b epigenetically activates neuroectodermal genes upon differentiation. In hESCs, nucleosome positioning is overall fuzzy. Nucleosomes are densely packed in promoters that are at a repressive chromatin state with prevalent H3K27me3 marks. Upon differentiation, nucleosome positioning becomes phased. H3K27me3 signals are greatly reduced. Concomitantly, Kat2b deposits H3K9ac signals that enhance Sox2 binding. H2A.Z-mediated nucleosomes eviction creates an open chromatin structure for transcription machinery assembly. As a result, neuroectodermal genes are activated for neuroectodermal commitment

**Figure 5** Kat2b acetylates the target sites of Sox2 and Pax6 for their binding. (a) Sox2 binds to distinct sites in hESCs and hNECs (heat map). Track view of Sox2-binding signals in the promoters of neuroectodermal TF Six6 and pluripotency TF Oct4. (b) Enrichment of co-factor motifs in hESC- and hNEC-specific Sox2-binding peak sequences. (c) Fold change of core histone modification signals in the vicinity of hNEC-specific Sox binding sites. (d) Fold change of Sox2 occupancy in  $\pm 2$  kb regions of CRRs' center. (e) Kat2b knockdown reduces H3K9ac levels and the binding of Sox2 and Pax6. Left: track view of H3K9ac signals in the regions spanning Sox2 or Pax6 binding peaks in the enhancer or promoter of three neuroectodermal TFs (Nrip1, Lhx2, Pax6) in the wild type cells. Right: qPCR validation of significantly decreased H3K9ac signals and the occupancy of Sox2 and Pax6 at their binding sites indicated by the dotted boxes in the left panel in hNECs after Kat2b knockdown. Error bars represent SEM. (\* $P < 0.05$ , \*\* $P < 0.01$ , Student's  $t$ -test)

**Knockdown of Kat2b in hESCs.** We knocked down Kat2b in hESCs by TALEN. CAG-GFP fragment was replaced by U6 promoter and shRNA sequence from AAV-CAGGS-EGFP (Addgene, Cambridge, MA, USA). The shRNA targeting sequences were: shKat2b#1 (5'-CGAACTCTAATCCTCACTCAT-3'), shKat2b#2 (5'-CCAGCCAGCTAGGCATCAAAA-3'), shKat2b#3 (5'-GGAAGCTGGATTAATTGA CAA-3'). Before electroporation, hESCs were cultured in ROCK-inhibitor Y-27632 (Millipore, Hayward, CA, USA, 688002) for 3 h. About  $1 \times 10^7$  cells were mixed with  $5 \mu\text{g}$  of each TALEN targeting AAVS1 locus.<sup>32</sup> And  $40 \mu\text{g}$  of AAVS1-SA-PURO-PA-U6 promoter-shRNA-PA donor after trypsin/EDTA solution treatment and then electroporated (Gene Pulser Xcell System, Bio-Rad: 250 V, 500  $\mu\text{F}$ , 0.4 cm cuvettes). Transfected cells were plated on MEF with ROCK-inhibitor for the first 24 h. Then we used  $1 \mu\text{g}/\text{ml}$  puromycin (Sigma, St. Louis, MO, USA, P8833) to select cells cultured with MEF-conditioned medium. Individual colonies were picked up after about 15 days and identified by PCR using primers for homologous recombination forward: 5'-CTTCCGCATTGGAGTCGCTTTA-3' and reverse: 5'-ACAGGAGGTGGGGTTAGAC-3' or wild type forward: 5'-CAGCCGG TCCTGGACTTTGTC-3' and reverse: 5'-AGCCGGAACCGCTCAA CTC-3'.

**Microarray analysis.** Gene expression microarray data of the two cell types were downloaded from ArrayExpress database (accession number E-MEXP-2426, Affymetrix HG-U133A).<sup>33</sup> The data contained 3 replicates. The CEL files were analyzed using the affy package within R/Bioconductor.<sup>34</sup> Robust multichip averaging (RMA) was used to correct for background, normalize and generate expression data.<sup>35</sup> The Limma (linear models for microarray data) package was then used to identify differentially expressed genes with the adjusted  $P$ -value  $< 0.05$  and expression fold change  $\geq 2$ .<sup>36</sup>

**Functional annotation of gene sets.** GO analysis for gene sets was performed using the tool PANTHER.<sup>37</sup> The tool Gene Set Enrichment Analysis (GSEA v2.1.0)<sup>24</sup> was used to identify a priori defined set of genes that share common biological function or regulation and show statistically significantly, concordantly different expression levels between two samples. GSEA was performed using the gene sets of gene ontology (C5) and user-defined HESC SPECIFIC MARKERS gene set that was collected from published paper.<sup>38</sup> Metric for ranking genes was set to Signal2Noise. Gene ontology gene sets were curated by MSigDB database v4.0. Other parameters used default values.

**Nucleosome prediction and analysis of nucleosome positioning dynamics.** Sequencing reads were aligned to *Homo sapiens* reference genome (hg19) using Bowtie-1.0.0 (ref. 39) with up to two mismatches. Only the uniquely mapped reads were retained for nucleosome prediction. Because only the border of mononucleosomal DNA was sequenced, each read was moved 73 bp interior to its end to represent nucleosome dyad. The peak-calling tool GeneTrack<sup>25</sup> was used to smooth the clustered distribution of reads and predict nucleosomes using an exclusion zone of 147 bp and sigma of 20 bp. Nucleosomes that were detected six or more times (that is, read count  $\geq 6$ ) were further analyzed, although patterns were identical when all nucleosomes were analyzed. Nucleosome fuzziness was calculated as the standard deviation of the coordinates of all reads defining the same nucleosome as described previously.<sup>40</sup> It measures how spread out a nucleosome position is. Each nucleosome was assigned to promoter, genic or intergenic regions depending on the location of the nucleosome midpoint.

Two nucleosomes from the two cell types are fixed if their midpoints locate at the same position. The nucleosome is gain or loss if the distance between two midpoints is  $\geq 107$  bp. That is, the two nucleosomes overlap less than 20 bp. The rest of nucleosomes shift upon differentiation.

**Nucleosome distribution profile around TSS.** The original composite distribution of nucleosome around TSS was calculated by aggregating nucleosomal read count at each distance relative to TSS as follows: each read represents a nucleosome by extending toward 3' end to a length of 147 bp. The midpoint of extended read defines the nucleosome position. We summed total read counts at each site within  $\pm 2$  kb of TSS. The nucleosomal read count was further normalized by the uniquely mapped total reads. We further binned the nucleosome occupancy by a 10-bp interval of nucleosome distance to TSS, and smoothed it with 5-bin moving average and 1-bin step size.

Nucleosome organization in the regions  $-0.3$  kb to  $+1$  kb of TSS was plotted as heat map by the tool seqMINER<sup>41</sup> without reads extension. Nucleosome positions predicted by GeneTrack was used as input with nucleosome width shrunk to 127 bp

for better visualization. Heat map for nucleosomes with H3K36me3 modifications was drawn in the same manner. Heat map for RNA Pol II was drawn in the similar way except that its peaks predicted by MACS-1.4.2<sup>(ref. 42)</sup> with default settings were used.

**Chromatin state in promoters.** HM ChIP-seq reads were aligned to *Homo sapiens* reference genome (hg19) as above nucleosomal reads. Similarly, only the uniquely mapped reads were retained for prediction of nucleosome with HM by GeneTrack using the same parameters as above nucleosome prediction. Occupancy of nucleosomes with HM were normalized as Reads Per Kilobase per Million mapped reads (RPKM). Nucleosomes with HM that had occupancy as RPKM  $> 1$  were retained for following analysis. Promoters ( $\pm 500$  bp of TSS) were categorized to eight classes (H3K4me3+, H3K27me3+, H3K9ac+, H3K4me3+/H3K27me3+, H3K4me3+/H3K9ac+, H3K9ac+/H3K27me3+, H3K4me3+/H3K27me3+/H3K9ac+, none) depending on the composition of nucleosomes with modifications located in promoters. For example, H3K4me3+ promoters contained only nucleosomes with H3K4me3, H3K4me3+/H3K27me3+ contained only both nucleosomes with H3K4me3 and nucleosomes with H3K27me3, and so on.

**Change in HMs in promoters.** Occupancy of a HM was measured as its read counts in promoters that were normalized as reads per kilobase per million mapped reads (RPKM). HM occupancy ratio of hNECs to hESCs was calculated, transformed to log<sub>2</sub>, and represented as heat map.

We further used ROC curve to identify which HM(s) were the best classifier for the significantly DEGs. In first case, upregulated DEGs were treated as actual positives, the rest were actual negatives. The fold changes of HM occupancy in promoters were calculated as above and sorted descendingly. Each fold change value was used as a threshold. For each threshold, the genes, whose fold changes of HM occupancy in promoters were greater than the threshold, were predictive positives. Otherwise, the genes are predictive negatives. Then, the true positive rate (TPR, that is, sensitivity) and false positive rate (FPR, that is,  $1 - \text{specificity}$ ) were calculated for each threshold. The ROC curve was thus TPR as a function of FPR. The ROC curve for downregulated DEGs was plotted in the same manner except that downregulated DEGs were treated as actual positives, the rest were actual negatives.

**Chromatin remodeling regions.** The regions with significant increase in H3K9ac upon differentiation were detected by the tool MACS-1.4.2<sup>(ref. 42)</sup> with a  $P$ -value cutoff of  $10^{-5}$  using hNECs as treatment sample and hESCs as background. The regions with significant decrease in H3K27me3 upon differentiation were detected similarly using as hESCs treatment sample and hNECs as background. We retained the regions with H3K9ac or H3K27me3 occupancy  $> 1$  RPKM. We combined these regions as chromatin remodeling regions. *De novo* motif discovery in these regions was done by the tool MEME<sup>43</sup> as follows: the DNA sequences of CRRs were retrieved as input to MEME for motif finding with  $E$ -value cutoff of 0.05 and motif width of 6–16 bp. The other parameters used default values. To compare the predicted motifs to the known ones, the TOMTOM motif comparison tool (Version 4.11.3) from the MEME suite was used to search a database of vertebrates (in vivo and in silico) using the default parameters ( $E$ -value  $< 10$ ).

**Profiles of Sox2 binding and motif finding in Sox2 peaks.** As above HM sequencing data analysis, only Sox2 reads uniquely mapped to *Homo sapiens* reference genome were retained for peak calling by the tool MACS-1.4.2 (ref. 42) with a  $P$ -value cutoff of  $10^{-4}$  and fold change of 10. The *de novo* motif finding in the Sox2 peak sequences was done by Cistrome Analysis Pipeline<sup>27</sup> as follows: we chose the SeqPos motif tool in the section Integrative Analysis from the Cistrome toolbox. The coordinates of Sox2 peaks in BED format was the input file. The '*de novo* motif search' motif database from the 'curated cistrome motif database' was used. The species '*Homo sapien* or *Mus musculus*' was used. The rest of parameters were default. The tool HOMER was applied to find enriched motifs of co-factors.<sup>44</sup> Only the motifs with the highest alignments to known transcription factors, nonredundant matrixes and non-repetitive sequences were retained. Sox2 occupancy in peaks was normalized as RPKM. Browser tracks for Sox2 occupancy were generated using the tool IGV 2.3.<sup>45</sup>

**Accession numbers.** The accession numbers for the MNase-seq data and the ChIP-seq data reported in this paper are GEO: GSE76084. The H2AZ ChIP-seq data was downloaded from GSM1003579. Pax6 binding peaks in hNECs was kindly provided by the author from the published work.<sup>46</sup>

**Conflict of Interest**

The authors declare no conflict of interest.

**Acknowledgements.** We thank Dr. Lawrence W Stanton for providing Pax6 binding peaks in hNECs. This work was supported by the National Key Research and Development Program of China (2016YFA0100400 to CJ), National Natural Science Foundation of China (91519309 and 31271373 to CJ).

- Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 1997; **389**: 251–260.
- Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M. Open chromatin in pluripotency and reprogramming. *Nat Rev Mol Cell Biol* 2011; **12**: 36–47.
- Li Z, Gadue P, Chen K, Jiao Y, Tuteja G, Schug J et al. Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. *Cell* 2012; **151**: 1608–1616.
- Ye Y, Gu L, Chen X, Shi J, Zhang X, Jiang C. Chromatin remodeling during the in vivo glial differentiation in early *Drosophila* embryos. *Sci Rep* 2016; **6**: 33422.
- Ye Y, Li M, Gu L, Chen X, Shi J, Zhang X et al. Chromatin remodeling during in vivo neural stem cells differentiating to neurons in early *Drosophila* embryos. *Cell Death Differ* 2017; **24**: 409–420.
- West JA, Cook A, Alver BH, Stadtfeld M, Deaton AM, Hochedlinger K et al. Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nat Commun* 2014; **5**: 4719.
- Kouzarides T. Chromatin modifications and their function. *Cell* 2007; **128**: 693–705.
- Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A et al. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* 2013; **153**: 1149–1163.
- Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* 2013; **153**: 1134–1148.
- Guenther MG, Frampton GM, Soldner F, Hockemeyer D, Mitalipova M, Jaenisch R et al. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* 2010; **7**: 249–257.
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007; **448**: 553–560.
- Zheng S, Sun M, Zhang K, Gu J, Guo Z, Tian S et al. Profiling post-translational modifications of histones in neural differentiation of embryonic stem cells using liquid chromatography-mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 2016; **1017–1018**: 36–44.
- Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD et al. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 2006; **442**: 91–95.
- Chen J, Chen X, Li M, Liu X, Gao Y, Kou X et al. Hierarchical Oct4 binding in concert with primed epigenetic rearrangements during somatic cell reprogramming. *Cell Rep* 2016; **14**: 1540–1554.
- Ziller MJ, Ediri R, Yaffe Y, Donaghey J, Pop R, Mallard W et al. Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature* 2015; **518**: 355–359.
- Pankratz MT, Li XJ, Lavaute TM, Lyons EA, Chen X, Zhang SC. Directed neural differentiation of human embryonic stem cells via an obligated primitive anterior stage. *Stem Cells* 2007; **25**: 1511–1520.
- 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**: 279–283.
- Zhang SC, Wernig M, Duncan ID, Brustle O, Thomson JA. In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nat Biotechnol* 2001; **19**: 1129–1133.
- Zhang XQ, Zhang SC. Differentiation of neural precursors and dopaminergic neurons from human embryonic stem cells. *Methods Mol Biol* 2010; **584**: 355–366.
- Zhang SC. Neural subtype specification from embryonic stem cells. *Brain Pathol* 2006; **16**: 132–142.
- Zhang X, Huang CT, Chen J, Pankratz MT, Xi J, Li J et al. Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* 2010; **7**: 90–100.
- Graham V, Khudyakov J, Ellis P, Pevny L. SOX2 functions to maintain neural progenitor identity. *Neuron* 2003; **39**: 749–765.
- Sarkar A, Hochedlinger K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* 2013; **12**: 15–30.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; **102**: 15545–15550.
- Albert I, Wachi S, Jiang C, Pugh BF. GeneTrack—a genomic data processing and visualization framework. *Bioinformatics* 2008; **24**: 1305–1306.
- Jiang C, Pugh BF. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 2009; **10**: 161–172.
- Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol* 2011; **12**: R83.
- Hawkins RD, Hon GC, Yang C, Antosiewicz-Bourget JE, Lee LK, Ngo QM et al. Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency. *Cell Res* 2011; **21**: 1393–1409.
- Flici H, Erkosar B, Komonyi O, Karatas OF, Laneve P, Giangrande A. Gcm/Glide-dependent conversion into glia depends on neural stem cell age, but not on division, triggering a chromatin signature that is conserved in vertebrate glia. *Development* 2011; **138**: 4167–4178.
- Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA et al. Super-enhancers in the control of cell identity and disease. *Cell* 2013; **155**: 934–947.
- Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 2013; **153**: 307–319.
- Sanjana NE, Cong L, Zhou Y, Cunniff MM, Feng G, Zhang F. A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* 2012; **7**: 171–192.
- Li XJ, Zhang X, Johnson MA, Wang ZB, Lavaute T, Zhang SC. Coordination of sonic hedgehog and Wnt signaling determines ventral and dorsal telencephalic neuron types from human embryonic stem cells. *Development* 2009; **136**: 4055–4063.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; **5**: R80.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; **4**: 249–264.
- Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004; **3**: Article 3.
- Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* 2013; **8**: 1551–1566.
- Cai J, Chen J, Liu Y, Miura T, Luo Y, Loring JF et al. Assessing self-renewal and differentiation in human embryonic stem cell lines. *Stem Cells* 2006; **24**: 516–530.
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009; **10**: R25.
- Shi J, Zheng M, Ye Y, Li M, Chen X, Hu X et al. *Drosophila* Brahma complex remodels nucleosome organizations in multiple aspects. *Nucleic Acids Res* 2014; **42**: 9730–9739.
- Ye T, Krebs AR, Choukallal MA, Keime C, Plewniak F, Davidson I et al. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res* 2011; **39**: e35.
- Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008; **9**: R137.
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 2009; **37**: W202–W208.
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010; **38**: 576–589.
- Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013; **14**: 178–192.
- Bhinge A, Poschmann J, Namboori SC, Tian X, Jia Hui Loh S, Traczyk A et al. MiR-135b is a direct PAX6 target and specifies human neuroectoderm by inhibiting TGF-beta/BMP signaling. *EMBO J* 2014; **33**: 1271–1283.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017

Supplementary Information accompanies this paper on Cell Death and Differentiation website (<http://www.nature.com/cdd>)