Front. Biol. 2016, 11(4): 311–322
DOI 10.1007/s11515-016-1411-5

RESEARCH ARTICLE

Locus- and cell type-specific epigenetic switching during cellular differentiation in mammals

Ying-Tao Zhao, Maria Fasolino, Zhaolan Zhou (✉)

Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

BACKGROUND: Epigenomic reconfiguration, including changes in DNA methylation and histone modifications, is crucial for the differentiation of embryonic stem cells (ESCs) into somatic cells. However, the extent to which the epigenome is reconfigured and the interplay between components of the epigenome during cellular differentiation remain poorly defined.
METHODS: We systematically analyzed and compared DNA methylation, various histone modification, and transcriptome profiles in ESCs with those of two distinct types of somatic cells from human and mouse.
RESULTS: We found that global DNA methylation levels are lower in somatic cells compared to ESCs in both species. We also found that 80% of regions with histone modification occupancy differ between human ESCs and the two human somatic cell types. Approximately 70% of the reconfigurations in DNA methylation and histone modifications are locus- and cell type-specific. Intriguingly, the loss of DNA methylation is accompanied by the gain of different histone modifications in a locus- and cell type-specific manner. Further examination of transcriptional changes associated with epigenetic reconfiguration at promoter regions revealed an epigenetic switching for gene regulation—a transition from stable gene silencing mediated by DNA methylation in ESCs to flexible gene repression facilitated by repressive histone modifications in somatic cells.
CONCLUSIONS: Our findings demonstrate that the epigenome is reconfigured in a locus- and cell type-specific manner and epigenetic switching is common during cellular differentiation in both human and mouse.

Keywords DNA methylation, histone modifications, epigenome, epigenetic switch, cellular differentiation

Introduction

Cellular differentiation is a precisely regulated process by which hundreds of morphologically and functionally distinct cell lineages are derived from the same origin, embryonic stem cells (ESCs). Given that ESCs and somatic cell types share an identical genome, a fundamental question is how lineage-specific gene expression patterns are established and maintained. Accumulating evidence supports that two epigenetic mechanisms, DNA methylation and histone modifications, are crucial in this process (Reik, 2007).

DNA methylation at the fifth carbon position of cytosine is an abundant covalent modification that primarily occurs in CpG dinucleotides (mCpGs), with 80% of CpGs being methylated in mammals. DNA methylation affects DNA–protein interactions and therefore influences transcription in various physiological processes, such as imprinting, X-inactivation, tumorigenesis, and embryogenesis (Jaenisch and Bird, 2003). In cellular differentiation, DNA methylation plays a critical role as the loss of enzymes responsible for establishing or removing methylation in embryos or ESCs results in skewed lineage specification and transdifferentiation (Jackson et al., 2004; Tsumura et al., 2006; Koh et al., 2011). Similar to DNA methylation, covalent modifications on histone tails play an imperative role in cellular differentiation and modulate transcription by altering DNA accessibility. Altering the activity of histone modifying enzymes, such as methyltransferases and acetyltransferases, diminishes the differentiation potential of ESCs (Margueron and Reinberg, 2011). Therefore, although histone modifications and DNA methylation are required for successful differentiation, the precise function of each in determining cell fate decisions is unknown.

Investigating the role of these epigenetic mechanisms has recently become possible with the construction of single base-pair resolution profiles of DNA methylation (methylomes) and genome-wide histone modification profiles in multiple

Received May 5, 2016; accepted June 21, 2016
Correspondence: Zhaolan Zhou
E-mail: zhaolan@mail.med.upenn.edu

cell types, tissues, and organisms (Lister and Ecker, 2009; Rivera and Ren, 2013). Comparison of methylomes across cellular differentiation has revealed both *de novo* methylation (Reik, 2007; Smith and Meissner, 2013) and *de novo* demethylation (Lister et al., 2009; Hon et al., 2012; Xie et al., 2013; Ziller et al., 2013). These findings raise the possibility that mCpG is more dynamic than previously thought. Indeed, literature supports that mCpG undergoes a demethylation process mediated by the TET family of proteins that oxidize mCpG to hydroxymethyl-CpG (Tahiliani et al., 2009). The acquisition of methylomes and histone modification profiles from multiple cell types also provides an opportunity to examine the interplay between these two epigenetic mechanisms. Increasing evidence supports that the loss of DNA methylation is coupled to the gain of histone modifications, such as H3K27me3 and H3K9me3 in tumorigenesis (Hon et al., 2012); H3K27me3, H3K9me3, and H3K4me1 in differentiation (Hawkins et al., 2010; Gifford et al., 2013); and H3K27ac and H3K4me1 in development (Lister et al., 2013). However, the interplay between these two epigenetic mechanisms has only been observed for a limited number of histone modifications, leaving the large scope of DNA-histone modification interactions to be explored.

In this study, we systematically integrated DNA CpG methylation, various histone modification, and transcriptome profiles in ESCs and two somatic cell types in both human and mouse. With the use of our novel unbiased, enrichment based statistical approach, we found that both DNA methylation and histone modifications are considerably altered during cellular differentiation. Approximately 70% of the reconfigurations in DNA methylation and histone modifications are cell type-dependent and locus-specific. Additionally, integrative analyses of these epigenomic reconfigurations revealed a widespread locus-specific switching from DNA methylation to histone modifications during cellular differentiation in both species. Furthermore, when these epigenetic changes at promoter regions were correlated with gene transcription, we found that the switching at promoter regions implicates a transition from stable gene silencing mediated by DNA methylation to flexible gene repression facilitated by histone modifications. Our study provides a comprehensive insight into the role of epigenomic reconfiguration in cellular differentiation.

Materials and methods

Bioinformatics analyses

All analyses in this study were processed in the MAC terminal window. Perl was used to do in-house programming. R (http://www.r-project.org/) and Bioconductor libraries (http://www.bioconductor.org) were used to do all statistical analyses. Integrative Genomics Viewer (IGV, version 2.2.4) (Robinson et al., 2011; Thorvaldsdóttir et al., 2013) was used as the local genome browser in this study. The first replicate of each histone modification (ChIP-seq) and gene expression (mRNA-seq) data set was used in the browser representation for all the figures (Table S1). Samtools version 0.1.18 (Li et al., 2009) was used to convert sam files to bam files and to sort the bam files by name. Sorted bam files were converted to tdf files by igvtools version 2.2.1 using parameters "-z 5-w 25-e 250" for ChIP-seq files and "-z 5-w 25-e 250 –strands read" for strand-specific mRNA-seq files. The track scales for ChIP-seq and mRNA-seq were normalized to the total number of uniquely mapped reads.

$$\begin{aligned}&\text{ChIP-seq\_scale}\\&=(\text{TDF\_scale}/\text{Total\_number\_uniquely\_mapped\_reads})\\&\quad*10\ 000\ 000\\&\text{mRNA-seq\_scale}\\&=(\text{TDF\_scale}/\text{Total\_number\_uniquely\_mapped\_reads})\\&\quad*20\ 000\ 000\end{aligned}$$

Obtaining the raw data and FASTQ files

The SRA raw data files for whole-genome bisulfite sequencing (WGBS), ChIP-seq, and mRNA-seq were downloaded from NCBI Sequence Read Archive database (http://www.ncbi.nlm.nih.gov/Traces/sra) under the accession numbers listed in Table S1. The SRA Toolkit version 2.1.7 (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view = software) was used to obtain the FASTQ files from the SRA files. The command used was "fastq-dump."

Genomic sequences and annotations

The reference genomic sequences used in this study were hg19 for the human and mm9 for the mouse. The genomic annotations, human GRCh37 and mouse NCBIM37 that were based on Ensembl release 66, were downloaded from Illumina iGenome database (http://cufflinks.cbcb.umd.edu/igenomes.html). The distal 5′ end of a gene was defined as the transcription start site (TSS). The exonic region was defined as all exons of a gene. The region from the 500 bp downstream of TSS to the distal 3′ end of a gene was defined as the gene body. Intergenic regions were downloaded from UCSC Genome Browser using Table Browser function (http://genome.ucsc.edu/cgi-bin/hgTables?command = start). Intergenic regions with at least 1 kb in length were included in subsequent analyses.

WGBS data mapping and methylation calling

The FASTQ raw data files were first processed by Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/

trim_galore/) to remove adaptor contamination, to trim lower-quality sequences, and to discard sequences shorter than 36 nucleotides. The Trim Galore parameters used were "-s 2-length 36." Bismark v0.7.4 (Krueger and Andrews, 2011) was used to map the WGBS data to the hg19 and the mm9 genomes. We used "-n 2-l 40" and default parameters in Bismark. We excluded reads that had three or more methylated cytosines in non-CpG contexts or had cytosine in the reads but thymine in the corresponding location of the reference genomes. For mapped reads from a library of the same PCR reaction with identical 5′ sites, the one with highest average phred quality score was kept and the others were discarded to get rid of possible PCR amplification artifacts. Libraries from multiple PCR reactions and from different replicates for each cell line were merged together. SNPs between C57BL/6 and 129 were downloaded from the Mouse Genome Project of the Sanger Institute (ftp://ftp-mouse.sanger.ac.uk/REL-1003/SNPs/20100310-all-snps.tab.gz). For the mouse methylomes, CpGs that overlapped with known SNPs were excluded from subsequent analyses. The methylation status of both strands of a CpG pair were merged together. CpGs with 10 or more read coverage were included in subsequent analyses. The methylation level for each CpG was calculated as follows:

$$mCpG = \frac{R_{C+} + R_{C-}}{(R_{C+} + R_{T+}) + (R_{C-} + R_{T-})}$$

where

R_{C+} is the number of reads with C mapped to the sense strand of a CpG, R_{T+} is the number of reads with T mapped to the sense strand of a CpG, R_{C-} is the number of reads with C mapped to the antisense strand of a CpG, R_{T-} is the number of reads with T mapped to the antisense strand of a CpG.

Genome-wide and locus-specific comparisons of DNA methylation levels

For genome-wide analysis, the two-tailed Fisher's exact test was used to compare the methylomes between two cell lines at single-CpG resolution. The *p*-value for each CpG was calculated as follows:

$$p\text{-value} = \frac{\binom{C_{ESC} + C_{Somatic}}{C_{ESC}}\binom{T_{ESC} + T_{Somatic}}{T_{ESC}}}{\binom{C_{ESC} + T_{ESC} + C_{Somatic} + T_{Somatic}}{C_{ESC} + T_{ESC}}}$$

$$= \frac{(C_{ESC} + C_{Somatic})!(T_{ESC} + T_{Somatic})!(C_{ESC} + T_{ESC})!(C_{Somatic} + T_{Somatic})!}{C_{ESC}!C_{Somatic}!T_{ESC}!T_{Somatic}!(C_{ESC} + T_{ESC} + C_{Somatic} + T_{Somatic})!}$$

where

$\binom{n}{k}$ is the binomial coefficient,

! is the factorial operator,

C_{ESC} is the number of methylated reads that covered this CpG in ESCs,

T_{ESC} is the number of non-methylated reads that covered this CpG in ESCs,

$C_{Somatic}$ is the number of methylated reads that covered this CpG in somatic cells,

$T_{Somatic}$ is the number of non-methylated reads that covered this CpG in somatic cells.

Raw *p*-values were adjusted using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995) to control the False Discovery Rate (FDR). CpGs with an adjusted p-value (FDR) less than 0.05 were defined as differentially methylated CpGs.

For locus-specific analysis, we developed an enrichment-based statistical approach to compare and identify differentially methylated genomic regions (Fig. 2A). We defined promoter regions using stringent criteria (Figure S2A) and analyzed them in parallel.

To determine the distribution of CpGs and mCpGs at promoter regions (Fig. 5A), the region 5 kb upstream and 1 kb downstream of TSS for each gene was dissected into 600 bins with 10 bp in length. For each bin, we calculated the total number of genomic CpGs, the ratio between the total number of mCpGs versus the total number of genomic CpGs, the ratio between the total number of diff-mCpGs and the total number of mCpGs, respectively.

ChIP-seq data mapping, peak calling, and sliding-window analysis

Bowtie version 0.12.7 (http://bowtie-bio.sourceforge.net/index.shtml) (Langmead et al., 2009) was used to map the FASTQ files of ChIP-seq data to the genomes. "-v 2-m 1-p 10" and default parameters were used. The mapping results for each replicate were merged together.

FindPeaks in Homer version 4.2 (http://biowhat.ucsd.edu/homer/ngs/) (Heinz et al., 2010) was used for peak calling using "-style histone" and default parameters. If the region 1 kb upstream and 1 kb downstream of the TSS for a gene overlapped with H3K27me3 or H3K9me3 peaks, we defined this gene as marked by that histone modification.

All replicates for each histone modification in each cell type were merged together. The total number of uniquely mapped reads was normalized to 40 million. For sliding window analysis, the human and mouse genomes were

partitioned into 1 kb sized windows, and the number of normalized reads for each window was calculated. The fold enrichment score for each histone modification at each window between ESCs and somatic cells was calculated as follows:

$$fold\ enrichment = \frac{R_{Somatic} + 1}{R_{ESC} + 1}$$

where $R_{Somatic}$ is the number of normalized reads in this window for somatic cells and R_{ESC} is the number of normalized reads in this window for ESCs.

Based on the fold enrichment score for each window, we classified all windows into three categories: those in which histone modifications are gained, lost, or remained. The gain windows were defined as those that have higher read numbers in somatic cells than in ESCs (fold enrichment score > 1), whereas the lose windows were defined as those that have lower read numbers in somatic cells than in ESCs (fold enrichment score < 1). We then obtained the top 2000 gain windows and the top 2000 lose windows based on their fold enrichment scores, and termed them as the highly dynamic regions. Only those dynamic regions with greater than or equal to 2 fold enrichment or depletion were included in subsequent analyses. The DNA methylation change associated with each highly dynamic region was calculated as follows:

$$methylation\ change = \frac{\sum_{i-1}^{n} S_i - \sum_{i-1}^{n} E_i}{n}$$

where n is the total number of CpGs with 10 or more read coverage in both cells in this region, S_i is the methylation level of the number i CpG in somatic cells, and E_i is the methylation level of the number i CpG in ESCs.

mRNA-seq data mapping and differential expression analysis

The mRNA-seq data for mouse frontal cortex were mapped to mouse NCBIM37/mm9 genomes by TopHat version 2.0.6 (Trapnell et al., 2009) using parameters "–bowtie1 –color –quals-p 5-g 1 –no-coverage-search –no-novel-juncs –library-type fr-secondstrand-G UCSC_mm9.gtf mm9_c." The other mRNA-seq data in FASTQ files were mapped to human GRCh37/hg19 or mouse NCBIM37/mm9 genomes by STAR for MAC version 2.2.1d (Dobin et al., 2013). We used STAR parameters "–outFilterMultimapNmax 1 –outFilterMismatchNmax 3" to only report uniquely mapped reads with no more than three mismatches.

The number of mapped reads in the exonic region of each gene was calculated using an in-house Perl program. All replicates for each cell type were calculated separately. Reads that overlapped with the exonic regions of multiple genes were excluded. The edgeR Bioconductor package (Robinson et al., 2010) was used to perform the differential expression analysis. The classic edgeR method, which is negative binomial model based exact test, was used. For differential expression analysis, we only included genes with a count per million (CPM) larger than 1 in at least two of the data sets. Genes with adjusted p-values less than 0.05 were defined as differentially expressed. Gene Ontology enrichment analysis was performed using DAVID v6.7 (http://david.abcc.ncifcrf.gov/) (Huang da et al., 2009a, 2009b).

Results

Global DNA hypomethylation in somatic cells

Given the essential role of DNA methylation in cellular differentiation, we first examined global DNA methylation in terminally differentiated somatic cells compared to ESCs using raw data from the NCBI GEO database (Table S1) (Lister et al., 2009; Stadler et al., 2011; Hon et al., 2012; Xie et al., 2012; Mann et al., 2013). Thus, we compared the methylomes from human H1 ESCs (hESC) and two distinct types of human somatic cells, IMR90 fetal lung fibroblasts (hFB) and mammary epithelial cells (hEP) (Figure S1A). To evaluate DNA methylation changes in a different species, we also compared the methylomes of mouse ESCs (mESC), mouse primary dermal fibroblasts (mFB), and mouse frontal cortex tissue (mFC) (Figure S1A).

We found that global mCpG levels are significantly lower in somatic cells compared to ESCs (Fig. 1). In human, 93% of the differentially methylated CpG sites (diff-mCpGs) exhibit a decrease in methylation in hFB compared to hESC and 89% of diff-mCpGs show a decrease in hEP (Fig. 1A–1B, two-tailed Fisher's exact test, Benjamini and Hochberg (B-H) correction). Global DNA hypomethylation was also found in mouse somatic cells (Fig. 1A–1B). Moreover, we found that hypomethylated diff-mCpGs are evenly distributed across all autosomes in both species, indicating widespread genome-wide DNA hypomethylation in somatic cells compared to ESCs (Figure S1C). Of note, global mCpG levels are not significantly different across somatic cell types (Figure S1B), suggesting that hypomethylation is found upon lineage commitment, rather than across cell types.

DNA hypomethylation is locus- and cell type-specific

Next, we investigated whether changes in DNA methylation occur at particular genomic features: intergenic regions, gene body, and promoter regions. To do so, we developed an enrichment based statistical approach for identifying genomic regions in which the ratio of diff-mCpGs is statistically higher than the ratio expected by chance (Fig. 2A). Thus, we were able to identify genomic regions that are differentially methylated between different cell types.

Consistent with our finding of global hypomethylation, a significantly larger portion of the differentially methylated regions is hypomethylated in somatic cells compared to ESCs

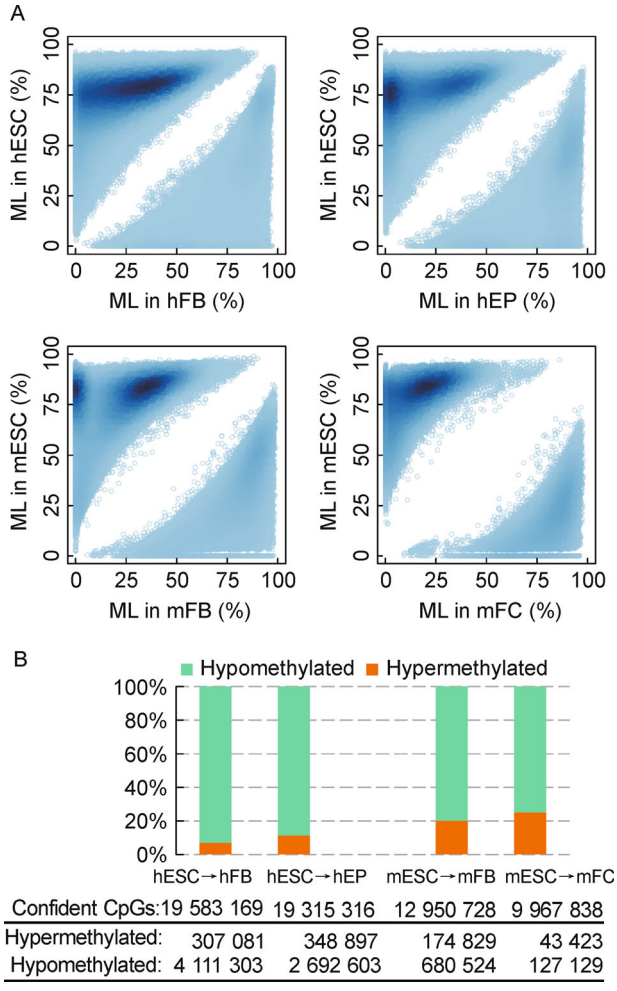

Figure 1 Global DNA hypomethylation in somatic cells in both human and mouse. (A) Methylation levels (ML) of differentially methylated CpGs (diff-mCpGs) between ESCs and two somatic cell types in human and mouse. Each circle represents a CpG. The deeper the shade of blue, the higher the point density (number of CpGs). (B) The percentage of diff-mCpGs (FDR < 0.05, two-tailed Fisher's exact test, B-H correction) that are hypomethylated or hypermethylated during cellular differentiation. Confident CpGs are those with 10 or more read coverage in both ESCs and somatic cells.

in both human and mouse (Figure S2A–S2C, FDR < 0.0001, Hypergeometric test, B-H correction). For example, > 90% of differentially methylated promoter, gene body, and intergenic regions are hypomethylated in hFB compared to hESC (Figure S2A–S2C). Notably, approximately 60%-80% of the identified differentially methylated genomic regions are unique to that specific somatic cell type regardless of being in the promoter, gene body, or intergenic region (Fig. 2B, Figure S2D). Genomic feature-specific alterations of DNA methylation are illustrated in the promoter regions of cell type-specific genes, including *LHX8*, *SH3TC1, COL6A3,* and *FOXI2* (Fig. 2C). Only 30% of the genomic regions identified in hFB and hEP share the same changes in DNA methylation when compared to hESC (Fig. 2B), such as the promoter regions of *TMEM173* and *TRIM4* (Fig. 2C). Furthermore, we found that the changes in DNA methylation are strictly confined in a locus-specific manner, as surrounding regions do not show similar alterations (Fig. 2C, Figure S2E– S2F). Together, these results demonstrate that the decrease in DNA methylation across cellular differentiation is lineage-specific and restricted to particular loci.

Locus- and cell type-specific alterations of histone modifications in human

Modification of histone tails, such as methylation, acetylation, ubiquitination or phosphorylation, is an evolutionarily conserved epigenetic mechanism crucial for gene regulation in many biological processes (Jenuwein and Allis, 2001; Berger, 2007). Characteristic histone modification features include H3K9me3 and H3K27me3 marks in transcriptionally silent loci and H3K4me3 and H3K27ac in transcriptionally active promoters (Turner, 2007). Moreover, genes that show transient transcriptional changes in response to environmental stimuli versus genes that are persistently transcribed are marked by distinct types of histone modifications (Métivier et al., 2003). Given the importance of histone modifications in regulating cellular processes, we next analyzed the extent to which histone modifications are altered in somatic cells compared to ESCs. We acquired all publicly available ChIP-seq data sets for histone modifications in hESC (26 data sets in total), hFB (26 data sets), and hEP (11 data sets) and compared the enrichment profiles for each modification across cell types (Table S1).

We found that 70%-80% of the enriched sites for each of the 26 histone modifications are altered in hFB compared to hESC and 60%-90% of the 11 histone modifications are altered between hEP and hESC, indicating a widespread epigenetic reorganization in human somatic cells (Fig. 3A). Comparison of histone modification profiles between the two somatic cell types revealed that alterations of histone modifications are predominately cell type-specific (Fig. 3B). For example, 70% of the identified H2A.Z sites that change during cellular differentiation are unique to each somatic cell type (Fig. 3B). Consistent with previous findings (Hawkins et al., 2010), we also found that repressive histone modifications, such as H3K27me3 and H3K9me3, are markedly enriched, including an increase in the total number of sites (Fig. 3A) as well as a significant expansion of occupancy for most of the enriched sites across cellular differentiation (Fig. 3C–3D, $p < 2.2 \times 10^{-16}$, Student's *t*-test). These results show that regions of histone modification occupancy change dramatically in a cell type-specific manner across cellular differentiation.

Epigenetic switching from DNA methylation to histone modifications

We next examined whether changes in histone modifications are associated with changes in DNA methylation across

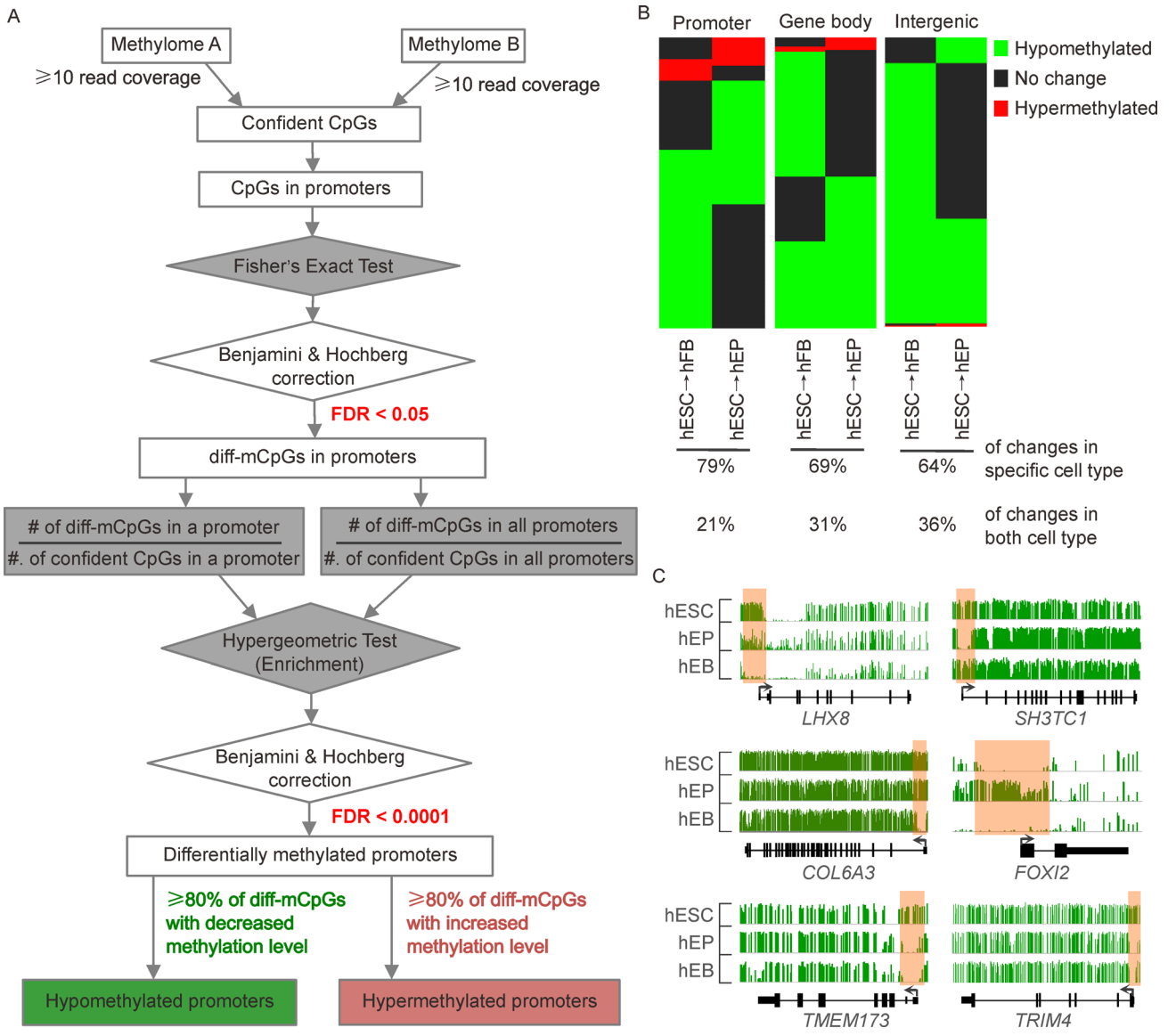

Figure 2 DNA hypomethylation during cellular differentiation is locus- and cell type-specific. (A) A schematic of the enrichment-based statistical approach used to identify differentially methylated genomic regions (e.g. promoters). Confident CpG, CpGs with 10 or more read coverage in both data sets; diff-mCpGs, differentially methylated CpGs. (B) Heat maps of differentially methylated genomic regions between hESC and the two human somatic cell types after clustering analysis. (C) Browser representation of methylation profiles of the genes with changes in methylation levels at their promoter regions during cellular differentiation. Each green vertical bar represents a CpG, and the height of the bar represents its methylation level (from 0 to 100%). Grey arrows indicate the transcriptional orientation of each gene. Regions marked by orange bars indicate the locus- and cell type-specific DNA methylation changes.

differentiation. To this end, we focused on genomic regions in somatic cells that exhibited the most dynamic changes in histone modifications relative to hESC. For each histone modification, these regions were defined as those with an enrichment score⩾2, and we limited our analyses to the top 4000 dynamic regions (2000 gain and 2000 loss).

We found that the majority of regions with increased histone modifications in somatic cells also exhibited decreased DNA methylation levels compared to hESC (Green boxes in Fig. 4A–4B). This correlation occurred for 85% of the histone modifications in hFB and 73% of modifications in hEP (Green boxes in Fig. 4A–4B). Moreover, we observed that many of the identified epigenetic switching from DNA methylation to histone modifications during cellular differentiation are specific to each cell type (Gray regions in Fig. 4D; Gray regions in Figure S3C). For example, two sites in the *RIN2* gene locus with specific loss of DNA methylation in hEP, but not in hFB, are accompanied by an increase in histone modifications in hEP, but not in hFB (Gray regions in Fig. 4D). Similarly, when we analyzed the relationship between DNA methylation and histone modifications in mouse mESC and mFC, we found that gains of

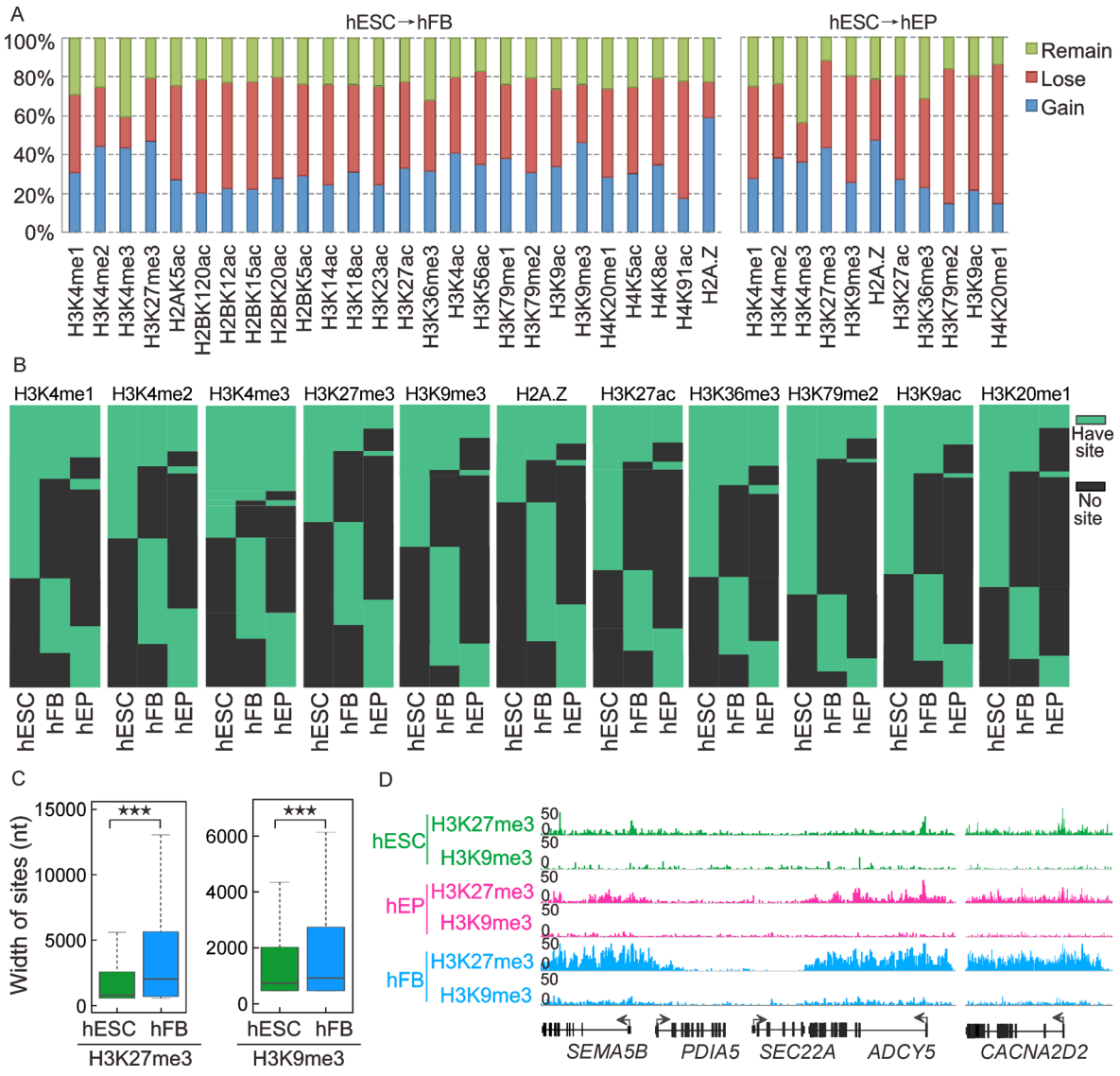

Figure 3 Changes in histone modifications during cellular differentiation are locus- and cell type-specific. (A) The percentage of enriched sites for each histone modification that were gained (gain), lost (lose), and remained (remain) during differentiation of hESC to hFB and hEP. (B) Heat maps of changes of enriched sites for 11 histone modifications during cellular differentiation in human after clustering analysis.(C) The width of H3K27me3 and H3K9me3 enriched sites in hESC and hFB. ★★★, p-value $< 2.2 \times 10^{-16}$, Student's t-test. (D) Browser representation of two regions with H3K27me3 and H3K9me3 expansion during cellular differentiation in human.

H3K4me3 and H3K27ac are associated with a decrease in DNA methylation, while the loss of H3K4me3 and H3K27ac are coupled with an increase in DNA methylation (Fig. 4C). This epigenetic switching is not found between regions exhibiting decreased histone modifications and DNA methylation for the majority of regions (Figure S3A–S3B). Together, these data demonstrate a cell type-dependent and locus-specific switching from DNA methylation to histone modifications over cellular differentiation in both species.

Epigenetic switching at promoters is associated with distinct gene regulation

We next set out to determine how switching from DNA methylation to histone modifications during cellular differentiation affects gene expression. We focused on promoter regions since changes in DNA methylation and histone modifications at promoter regions are correlated with changes in transcription relative to other genomic features (Ball et al., 2009). We found that proximal promoters, regions 0.5kb upstream and downstream of transcriptional start sites (TSSs), have a higher CpG density and are more dynamically methylated than distal promoter regions, defined as regions from 0.5 to 5kb upstream and 0.5-1kb downstream of TSSs (Fig. 5A, Figure S4A). Given this finding, we identified differentially methylated proximal promoters over differentiation with the use of our enrichment-based statistical approach (Fig. 2A). Compared to ESCs, we found that the

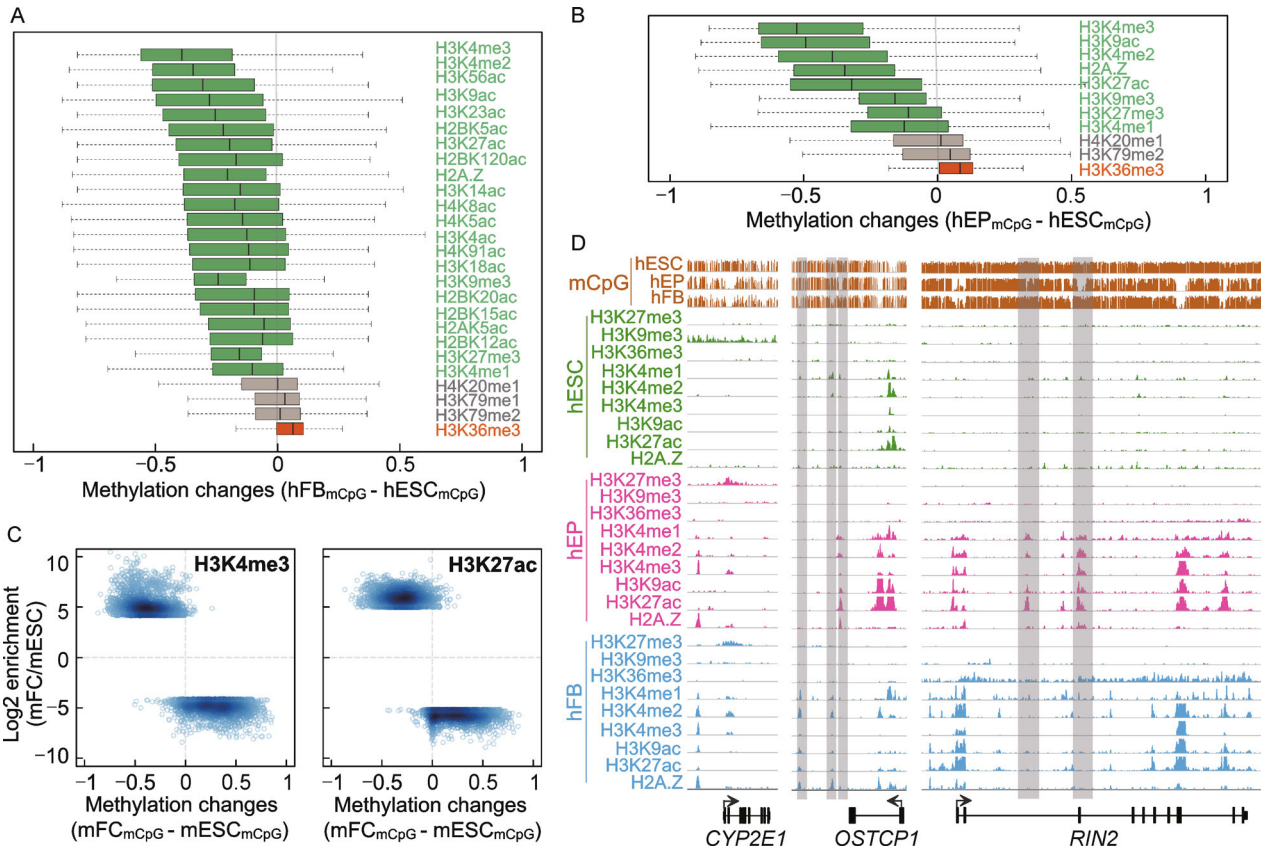

Figure 4 A switching from DNA methylation to histone modifications during cellular differentiation. (A) DNA methylation changes in the highly dynamic regions that gain histone modifications during the differentiation of hESC to hFB. Green regions, histone modifications that are associated with a decrease in DNA methylation; Gray regions, histone modifications that are not associated with a change in DNA methylation; Red regions, histone modifications that are associated with an increase in DNA methylation. (B) DNA methylation changes in the highly dynamic gain regions during the differentiation of hESC to hEP. (C) DNA methylation changes in the highly dynamic regions that gain or lose H3K4me3 and H3K27ac during the differentiation of mESC to mFC. Enrichment score for each region was calculated by dividing the number of normalized reads aligned in a particular region in mFC by that in mESC. (D) Browser representation of three regions that exhibit switching from DNA methylation to histone modifications during cellular differentiation in human. Regions marked by gray bars indicate the cell type-specific switching. The track scale for all histone modifications is from 0 to 50 normalized reads.

majority of differentially methylated promoters were hypomethylated: 88% of the promoters in hFB, 82% of the promoters in hEP, 77% of the promoters in mFB, and 65% of the promoters in mFC (Table S2–S5). Therefore, we focused on hypomethylated promoters for subsequent analyses. Interestingly, gene ontology (GO) analysis of genes with hypomethylated proximal promoters revealed significant enrichment in processes characteristic of each somatic cell type as well as pathways responsive to environmental signals (Figure S4B, Fisher′s exact test, B-H correction). For example, many of the significant GO terms for hFB were related to cell adhesion, which is an essential characteristic of hFB, and injury response, a pathway highly dependent on external cues (Figure S4B). These findings suggest that hypomethylation at promoter regions in somatic cells is important for defining cellular identity and rendering them responsive to environmental cues.

Examination of transcriptional changes of genes with hypomethylated promoters revealed that 25% of genes showed an increase in expression in hFB compared to hESC (Red regions in Fig. 5B). This is consistent with the prevailing model that a decrease in DNA methylation at promoters correlates with an increase in gene expression (Ball et al., 2009). However, 65% of genes identified in hFB did not exhibit changes in transcription despite a significant decrease in DNA methylation at their promoters (Black regions in Fig. 5B, FDR < 0.0001, Hypergeometric test, B-H correction), with 78% of these genes are transcriptionally silent in both hESC and hFB (RPKM < 1). Alternative promoter usage (Maniatis and Reed, 2002) may account for some of the genes in which promoter hypomethylation is associated with a decrease in gene expression in hFB (Green regions in Fig. 5B, Figure S5A). Similar percentages of genes with hypomethylated promoters show an increase (34%), no change (50%), or decrease (16%) in gene expression in hEP compared to hESC (Fig. 5B). In mice, promoter hypomethylation had similar

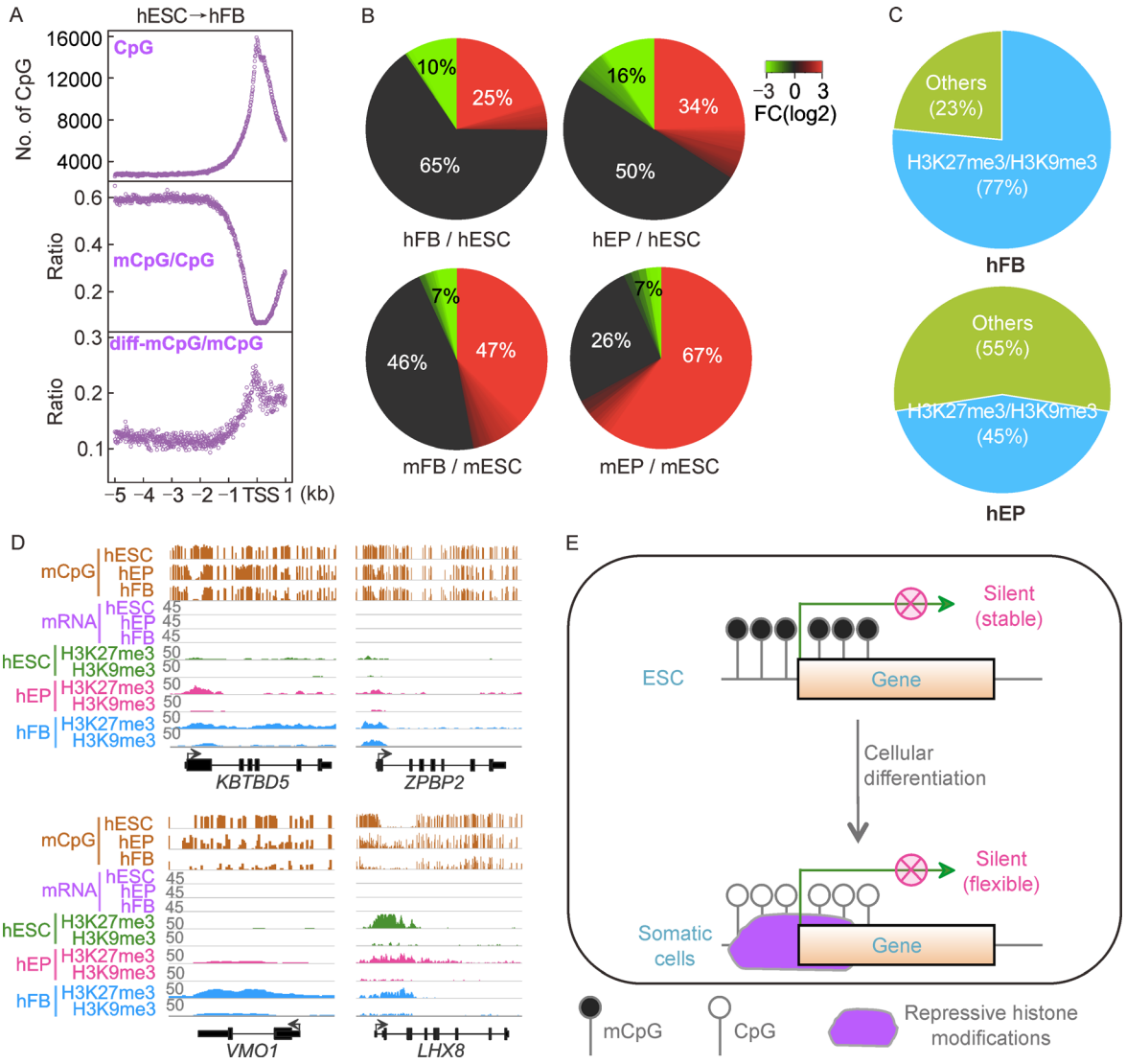

Figure 5 Switching from DNA methylation to repressive histone modifications at promoter regions uncovers a change of gene regulation from ESCs to somatic cells. (A) Scatter plots of the number of genomic CpGs, the ratio of mCpG, and the ratio of diif-mCpG in 10 bp bins within the region 5 kb upstream and 1 kb downstream of known transcription start site (TSS). (B) Heat map-pie charts of the expression changes of the genes with hypomethylated promoters. FC, fold change, gene expression levels in somatic cell types divided by that in ESCs. (C) Pie charts of percentages of genes, which are hypomethylated at the promoter regions but with no expression changes during cellular differentiation, marked by H3K27me3 or H3K9me3 or both in somatic cells. (D) Browser representation of four genes with epigenomic reconfiguration at their promoter regions. These genes were silent in both hESC and somatic cell types. The track scale for all mRNA-seq is from 0 to 45 normalized reads. The track scale for all histone modifications is from 0 to 50 normalized reads. (E) A model depicting gene regulation changes during cellular differentiation.

effects on gene expression with the majority of genes with hypomethylated promoters either showing an increase in gene expression, 47% in mFB and 67% in mFC, or no change in gene expression, 46% in mFB and 26% in mFC, while only a small percentage exhibited a decrease in gene expression, 7% in mFB and 7% in mFC) (Fig. 5B). These data suggest that promoter hypomethylation during cellular differentiation is not only important for increasing lineage specific transcripts, but also for flexible gene repression facilitated by repressive histone modifications in somatic cells.

To gain an insight into the mechanism by which genes with hypomethylated promoters remain transcriptionally silent in human somatic cells, we examined the profiles of two repressive histone modifications, H3K27me3 and H3K9me3, at the promoter regions of these silent genes in hFB and hEP. We found that, 77% of these genes in hFB and 45% in hEP are marked by one or both of H3K27me3 and H3K9me3 at their promoter regions (Fig. 5C–5D, Figure S5B). Furthermore, these repressive histone modifications at the promoter regions are either established *de novo* during cellular differentiation, such as at the promoter regions of *KBTBD5*, *ZPBP2*, and *VMO1* (Fig. 5D), or inherited from hESC, such as at the

promoter region of *LHX8* (Fig. 5D). Given that DNA methylation is known to be a stable repressive mechanism (Bird, 2002; Jones, 2012), the epigenetic switching at the promoter regions of these genes during cellular differentiation may allow them to be dynamically expressed in response to external cues in somatic cells (Fig. 5E). Together, these results uncover a potential role for epigenomic reconfiguration in facilitating gene regulation, by which genes stably silenced by DNA methylation in ESCs switch to flexible repression mediated by repressive histone marks in somatic cells.

Discussion

Two longstanding questions in cellular differentiation are how ESCs maintain pluripotency and how they give rise to a multitude of diverse types of somatic cells. Given that ESCs and somatic cells share identical genomic sequences, the establishment and maintenance of each somatic cell identity has been attributed to epigenetic modifications. This has led to an appealing model in which a flexible epigenome in ESCs allows for a multitude of genes to be transiently activated or repressed to support pluripotency.

In support of this model, studies have found an increase in DNA methylation at the critically important pluripotency gene *Oct4* (Ben-Shushan et al., 1993; Deb-Rinker et al., 2005). Additionally, the promoters of key developmental regulators are enriched with histone modifications, such as H3K27me3 or H3K4me3, rather than more stable DNA methylation in ESCs compared to somatic cells (Azuara et al., 2006; Bernstein et al., 2006; Pan et al., 2007). A limitation of these studies is the biased focus on loci involved in pluripotency and development. Therefore, to fully understand the role of the epigenome in differentiation, we took advantage of publicly available data sets from ESCs and distinct types of somatic cells in both human and mouse and performed an unbiased, genome-wide analysis of DNA methylation, histone modifications, and transcription across differentiation. Our data provides evidence against the prevailing model in which ESCs have a more flexible epigenome than somatic cells and instead supports that somatic cells have a more adaptable epigenetic landscape.

By analyzing two components of the epigenome, DNA methylation and histone modifications, across cellular differentiation in two somatic cell types in human and mouse, we found extensive, cell type-specific changes. Global analysis of DNA methylation revealed that millions of CpGs are differentially methylated across differentiation, with 90% of these differentially methylated CpGs being hypomethylated. Additionally, we found that 70% of these differentially methylated CpGs across differentiation are unique to each somatic cell type, which is in agreement with a recent finding of tissue-specific methylation patterns in mouse (Hon et al., 2013). Strikingly, we found that genomic regions that are hypomethylated in somatic cells are accompanied by a gain of 22 different histone modifications in a cell type-specific manner. This suggests that the switching from DNA methylation to histone modifications is an important epigenetic event for differentiation and lineage specification.

The association between DNA hypomethylation and the gain of histone modifications has been suggested for only a few histone modifications, such as H3K27me3, H3K9me3, and H3K4me1 (Lister et al., 2009; Hawkins et al., 2010; Gifford et al., 2013). Our results have identified 18 new histone modifications that correlate with hypomethylation. We found that genes repressed in both ESCs and somatic cells are held silent in ESCs by DNA methylation, a more permanent epigenetic modification, whereas in somatic cells genes are repressed by histone modifications, a more transient modification. This switching from DNA methylation to histone modifications at repressed genes across differentiation supports that the epigenome is more flexible in terminally differentiated cells than in ESCs.

Our finding that epigenomic reconfiguration occurs in a cell type-dependent and locus-specific manner raises the importance of profiling epigenomic information in homogeneous cell populations. Although cellular heterogeneity has been difficult to address, especially for complex tissues such as the brain, advanced technology in combination with newly designed approaches should make this feasible in the near future. The data processing pipelines developed in this study can allow us to address the extent of reconfiguration, and interplay with histone modifications, including recently discovered hmCpG and non-mCpG, once genomic data for these modifications in different cell types become available.

Abbreviations

ESCs: embryonic stem cells;
mCpGs: methylated CpGs;
hESC: human H1 ESCs;
hFB: IMR90 fetal lung fibroblasts;
hEP: mammary epithelial cells;
mESC: mouse ESCs;
mFB: mouse primary dermal fibroblasts;
mFC: mouse frontal cortex;
diff-mCpGs: differentially methylated CpG sites;
B-H: Benjamini and Hochberg;
TSSs: transcriptional start sites;
GO: gene ontology;
IGV: Integrative Genomics Viewer;
WGBS: whole-genome bisulfite sequencing;
FDR: False Discovery Rate;
CPM: count per million.

Acknowledgements

We thank members of the Zhou laboratory for discussions and comments

and the ENCODE Consortium and NIH Roadmap Epigenomics Mapping Consortium for availability of data sets. This work was funded by grants from the National Institutes of Health (R01MH091850 to Z. Z.), and Z.Z. is a Pew scholar in biomedical sciences.

Compliance with ethics guidelines

Ying-Tao Zhao, Maria Fasolino and Zhaolan Zhou declares that they have no conflict of interest. This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen H F, John R M, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher A G (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol, 8(5): 532–538

Ball M P, Li J B, Gao Y, Lee J H, LeProust E M, Park I H, Xie B, Daley G Q, Church G M (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol, 27(4): 361–368

Ben-Shushan E, Pikarsky E, Klar A, Bergman Y (1993). Extinction of Oct-3/4 gene expression in embryonal carcinoma x fibroblast somatic cell hybrids is accompanied by changes in the methylation status, chromatin structure, and transcriptional activity of the Oct-3/4 upstream region. Mol Cell Biol, 13(2): 891–901

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate- a practical and powerful approach to multiple testing. J Roy Stat Soc B Met, 57: 289–300

Berger S L (2007). The complex language of chromatin regulation during transcription. Nature, 447(7143): 407–412

Bernstein B E, Mikkelsen T S, Xie X, Kamal M, Huebert D J, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber S L, Lander E S (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell, 125 (2): 315–326

Bird A (2002). DNA methylation patterns and epigenetic memory. Genes Dev, 16(1): 6–21

Deb-Rinker P, Ly D, Jezierski A, Sikorska M, Walker P R (2005). Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. J Biol Chem, 280(8): 6257–6260

Dobin A, Davis C A, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras T R (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1): 15–21

Gifford C A, Ziller M J, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek A K, Kelley D R, Shishkin A A, Issner R, Zhang X, Coyne M, Fostel J L, Holmes L, Meldrim J, Guttman M, Epstein C, Park H, Kohlbacher O, Rinn J, Gnirke A, Lander E S, Bernstein B E, Meissner A (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell, 153(5): 1149–1163

Hawkins R D, Hon G C, Lee L K, Ngo Q, Lister R, Pelizzola M, Edsall L E, Kuan S, Luu Y, Klugman S, Antosiewicz-Bourget J, Ye Z, Espinoza C, Agarwahl S, Shen L, Ruotti V, Wang W, Stewart R, Thomson J A, Ecker J R, Ren B (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell, 6(5): 479–491

Heinz S, Benner C, Spann N, Bertolino E, Lin Y C, Laslo P, Cheng J X, Murre C, Singh H, Glass C K (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell, 38(4): 576–589

Hon G C, Hawkins R D, Caballero O L, Lo C, Lister R, Pelizzola M, Valsesia A, Ye Z, Kuan S, Edsall L E, Camargo A A, Stevenson B J, Ecker J R, Bafna V, Strausberg R L, Simpson A J, Ren B (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res, 22(2): 246–258

Hon G C, Rajagopal N, Shen Y, McCleary D F, Yue F, Dang M D, Ren B (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat Genet, 45(10): 1198–1206

Huang W, Sherman B T, Lempicki R A (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res, 37(1): 1–13

Huang W, Sherman B T, Lempicki R A (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc, 4(1): 44–57

Jackson M, Krassowska A, Gilbert N, Chevassut T, Forrester L, Ansell J, Ramsahoye B (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol Cell Biol, 24(20): 8862–8871

Jaenisch R, Bird A (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet, 33(3s Suppl): 245–254

Jenuwein T, Allis C D (2001). Translating the histone code. Science, 293 (5532): 1074–1080

Jones P A (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet, 13(7): 484–492

Koh K P, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer C A, Mostoslavsky G, Lahesmaa R, Orkin S H, Rodig S J, Daley G Q, Rao A (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell, 8(2): 200–213

Krueger F, Andrews S R (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics, 27 (11): 1571–1572

Langmead B, Trapnell C, Pop M, Salzberg S L (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 10(3): R25

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, the 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics, 25(16): 2078–2079

Lister R, Ecker J R (2009). Finding the fifth base: genome-wide sequencing of cytosine methylation. Genome Res, 19(6): 959–966

Lister R, Mukamel E A, Nery J R, Urich M, Puddifoot C A, Johnson N D, Lucero J, Huang Y, Dwork A J, Schultz M D, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu J C, Rao A, Esteller M, He C, Haghighi F G, Sejnowski T J, Behrens M M, Ecker J R (2013). Global epigenomic reconfiguration during mammalian brain development. Science, 341(6146): 1237905

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

Maniatis T, Reed R (2002). An extensive network of coupling among gene expression machines. Nature, 416(6880): 499–506

Mann I K, Chatterjee R, Zhao J, He X, Weirauch M T, Hughes T R, Vinson C (2013). CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. Genome Res, 23(6): 988–997

Margueron R, Reinberg D (2011). The Polycomb complex PRC2 and its mark in life. Nature, 469(7330): 343–349

Métivier R, Penot G, Hübner M R, Reid G, Brand H, Kos M, Gannon F (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. Cell, 115(6): 751–763

Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir G A, Stewart R, Thomson J A (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell, 1(3): 299–312

Reik W (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature, 447(7143): 425–432

Rivera C M, Ren B (2013). Mapping human epigenomes. Cell, 155(1): 39–55

Robinson J T, Thorvaldsdóttir H, Winckler W, Guttman M, Lander E S, Getz G, Mesirov J P (2011). Integrative genomics viewer. Nat Biotechnol, 29(1): 24–26

Robinson M D, McCarthy D J, Smyth G K (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1): 139–140

Smith Z D, Meissner A (2013). DNA methylation: roles in mammalian development. Nat Rev Genet, 14(3): 204–220

Stadler M B, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley E J, Gaidatzis D, Tiwari V K, Schübeler D (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature, 480(7378): 490–495

Tahiliani M, Koh K P, Shen Y, Pastor W A, Bandukwala H, Brudno Y, Agarwal S, Iyer L M, Liu D R, Aravind L, Rao A (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science, 324(5929): 930–935

Thorvaldsdóttir H, Robinson J T, Mesirov J P (2013). Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform, 14(2): 178–192

Trapnell C, Pachter L, Salzberg S L (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 25(9): 1105–1111

Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda H R, Nakayama J, Okano M (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells, 11(7): 805–814

Turner B M (2007). Defining an epigenetic code. Nat Cell Biol, 9(1): 2–6

Xie W, Barr C L, Kim A, Yue F, Lee A Y, Eubanks J, Dempster E L, Ren B (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell, 148(4): 816–831

Xie W, Schultz M D, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker J W, Tian S, Hawkins R D, Leung D, Yang H, Wang T, Lee A Y, Swanson S A, Zhang J, Zhu Y, Kim A, Nery J R, Urich M A, Kuan S, Yen C A, Klugman S, Yu P, Suknuntha K, Propson N E, Chen H, Edsall L E, Wagner U, Li Y, Ye Z, Kulkarni A, Xuan Z, Chung W Y, Chi N C, Antosiewicz-Bourget J E, Slukvin I, Stewart R, Zhang M Q, Wang W, Thomson J A, Ecker J R, Ren B (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell, 153(5): 1134–1148

Ziller M J, Gu H, Müller F, Donaghey J, Tsai L T, Kohlbacher O, De Jager P L, Rosen E D, Bennett D A, Bernstein B E, Gnirke A, Meissner A (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature, 500(7463): 477–481