Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press



Spatiotemporal clustering of epigenome reveals rules of dynamic gene regulation

Pengfei Yu, Shu Xiao, Xiaoyun Xin, et al.

Genome Res. published online October 2, 2012
Access the most recent version at doi:10.1101/gr.144949.112

P<P Published online October 2, 2012 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**



To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

© 2012, Published by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Spatiotemporal clustering of epigenome reveals rules of dynamic gene regulation

Authors:

Pengfei Yu[1,3,6], Shu Xiao[2,3,6], Xiaoyun Xin[5,7], Chun-Xiao Song[4,7], Wei Huang[2,3,7], Darina McDee[1], Tetsuya Tanaka[3], Ting Wang[5], Chuan He[4], Sheng Zhong[2,3,*]

Author affiliations:

[1] School of Molecular and Cellular Biology, [2] Department of Bioengineering, [3] Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA, 61801

[4] Department of Chemistry, University of Chicago, Chicago, IL, USA 60637

[5] Department of Genetics, Washington University in St. Louis, St. Louis, MO, USA 63108

[6, 7] These authors contributed equally.

* Correspondence should be addressed to: szhong@illinois.edu

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Abstract

Spatial organization of different epigenomic marks was used to infer functions of the epigenome. It remains unclear what can be learned from the temporal changes of the epigenome. Here, we developed a probabilistic model to cluster genomic sequences based on the similarity of temporal changes of multiple epigenomic marks during a cellular differentiation process. We differentiated mouse embryonic stem (ES) cells into mesendoderm cells. At three time points during this differentiation process, we used high-throughput sequencing to measure 7 histone modifications and variants including H3K4me1/2/3, H3K27ac, H3K27me3, H3K36me3, and H2A.Z, 2 DNA modifications including 5-mC and 5-hmC, and transcribed mRNAs and non-coding RNAs (ncRNAs). Genomic sequences were clustered based on the spatiotemporal epigenomic information. These clusters not only clearly distinguished gene bodies, promoters, and enhancers, but also were predictive of bidirectional promoters, miRNA promoters, and piRNAs. This suggests specific epigenomic patterns exist on piRNA genes much earlier than germ cell development. Temporal changes of H3K4me2, unmethylated CpG, and H2A.Z were predictive of 5-hmC changes, suggesting unmethylated CpG and H3K4me2 as potential upstream signals guiding TETs to specific sequences. Several rules on combinatorial epigenomic changes and their effects on mRNA expression and ncRNA expression were derived, including a simple rule governing the relationship between 5-hmC and gene expression levels. A *Sox17* enhancer containing a FOXA2 binding site and a *Foxa2* enhancer containing a SOX17 binding site were identified, suggesting a positive feedback loop between the two mesendoderm transcription factors. These data illustrate the power of using epigenome dynamics to investigate regulatory functions.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Introduction

An epigenome consists of chemical modifications and protein variations to the DNA and histones, and some of these modifications and variations can be passed down to an organism's offspring (Bernstein et al. 2007). Epigenomes are dynamic, and epigenetic modifications are associated with changes in gene expression (Creyghton et al. 2010; Hawkins et al. 2011). Thus, the epigenome adds an extra layer of information onto the genomic sequence and enables a genome to dynamically orchestrate gene expression in different cell types (Karlic et al. 2010; Maunakea et al. 2010). It is argued that organismal development can be viewed as a progression of epigenomic states (Bernstein et al. 2007; Hawkins et al. 2010). To gain mechanistic support for this view, a number of challenges have to be addressed. First, when presented with genome-wide distributions of epigenetic modifications at multiple time points during a developmental or differentiation process, how can we find the genomic (cis-) regulatory sequences that regulate gene expression? What are the combinatorial functions of epigenetic modifications and regulatory sequences? Here, we present experimental data and a probabilistic model that utilizes the temporal changes of the epigenome to annotate the regulatory sequences. This approach classifies regulatory sequences by their temporal epigenomic patterns, and thus it can identify subclasses of cis- regulatory sequences with different regulatory functions.

Two types of associations were observed between the epigenome and gene expression. First, in a given cell type, the transcription levels of different genes are associated with the epigenomic modifications in the genomic neighborhoods of these genes. In other words, without changing cell types, epigenomic modifications at different chromosomal locations are indicative of the relative abundance of RNAs transcribed from these locations (Karlic et al. 2010) (spatial correlation (S), Figure 1A). Second, for a given gene, the temporal change in its expression during a developmental or differentiation process is associated with temporal epigenomic changes (Rada-Iglesias et al. 2011) (temporal correlation (T), Figure 1A). The first type of association facilitated the use of invariant epigenomic signatures in a static cellular condition to annotate genomic features (Ernst and Kellis 2010). However, genes are dynamically regulated in nearly all biological processes. It is important to incorporate the dynamic aspect of gene regulation into the annotation of the regulatory sequences. Here, we jointly model the position effect and the temporal effect of the epigenome, thus achieving *ab initio* identification and functional annotation of regulatory sequences.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

The regulatory functions of a number of epigenetic modifications remain elusive. A case in point is DNA hydroxymethylation (Wyatt and Cohen 1952). Methylated cytosine (5-mC) can be converted to an oxidized form 5-hydroxymethylcytosine (5-hmC) by a family of Ten-eleven translocation (TET) proteins (Tahiliani et al. 2009; Ito et al. 2010). In embryonic stem (ES) cells, 5-hmC is enriched in gene bodies of actively transcribed genes (Wu et al. 2011a), promoters of inactive genes (Ficz et al. 2011; Pastor et al. 2011; Wu et al. 2011a), and active enhancers (Stroud et al. 2011). These seemingly conflicting data are thought-provoking for analyzing epigenetic modifications in a combinatorial manner, such that the function of each modification is investigated in the context of other modifications as well as the underlying genomic sequence. By modeling the co-appearance of different epigenetic modifications in each cell type, a pioneering method demonstrated the power in predicting different genomic features, including enhancers and genes (Ernst and Kellis 2010). However, epigenomic co-appearance in static cell types does not reveal all epigenetic mechanisms of gene regulation. Two major questions remain unsolved. First, what are the upstream signals that guide specific epigenomic modifications, such as 5-hmC, to appear in specific genomic regions? Second, the regulatory functions for several epigenomic marks including 5-hmC and H2A.Z remain elusive. New ideas for combinatorial epigenomic analyses beyond the co-appearance in static cell types are needed. A natural extension in this direction is to utilize a dynamic process in which both the epigenome and the transcriptome have changes. Ideally, we need some methods that can capture combinatorial patterns of temporal epigenomic changes and correlate them with gene expression changes.

A major difficulty in analyzing epigenomic dynamics lies in the asynchronous nature of epigenomic changes in different genomic regions. Suppose a type of epigenomic change, for example the induction of H3K4me1 and 5-hmC, is a recurring pattern shared by many genomic regions. Such a pattern can be difficult to find because different genomic regions can accumulate either modification at different times. Furthermore, the corresponding changes in gene expression are not synchronized either, making it difficult to associate epigenomic dynamics with gene expression changes. To reveal the hidden rules of epigenomic dynamics and gene expression, we developed a spatiotemporal clustering model. This model clusters genomic regions by shared epigenomic changes but does not require the changes to be synchronized among a cluster of genomic regions. This was achieved by allowing each region to have its own time-specific *epigenomic states* and then integrating out the time of transition between the epigenomic states in the clustering model.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

To investigate the functions of epigenetic modifications in a dynamic process, we differentiated mouse embryonic stem (ES) cells into mesendoderm cells (Yasunaga et al. 2005), the common precursor of mesoderm and endoderm. At 3 time points during this differentiation process, we mapped the genomic distributions of 9 epigenetic modifications, including DNA methylation (5-mC), hydroxymethylation (5-hmC), histone variant H2A.Z, and histone modifications H3K4me1, H3K4me3, H3K27ac, H3K27me3, and H3K36me3. At the same time points, we also assayed the expression of small non-coding RNAs (ncRNAs) and total RNAs (Table S1). Our model-based analysis of these temporal data revealed several fundamental properties of epigenome dynamics, characterizing regulatory roles for functionally elusive epigenomic modifications. As an analogy to the sequence "rules" of gene regulation (Buchler et al. 2003; Beer and Tavazoie 2004), these discoveries may provide epigenomic "rules" of gene regulation.

Results

A model for genome annotation using temporal epigenomic data

We developed a probabilistic model to annotate the genome using temporal epigenomic data. Two main features of this model include explicit treatment of combinatorial epigenomic changes and detecting similar but asynchronous epigenomic changes in different genomic segments.

As input data to the model, the genome is represented as consecutive genomic segments, with a typical segment size of 200 nucleotides (nt) (see Figure S5 for the impact of segment sizes). Each segment is associated with the time-specific intensities of a set of epigenetic modifications. The model clusters the genomic segments, such that each cluster shares a similar combination of epigenetic modifications as well as their temporal changes. We call the combination of epigenetic modifications shared by a cluster of genomic segments at a given time an *epigenomic state*. Each cluster represents a time-series of related epigenomic states. Essentially this model assigns epigenomic states based on time-series epigenomic data.

We call this model Genomic Annotation using Temporal Epigenomic data (GATE). GATE is a hierarchical model (Pearl 1985) with two layers (Figure 1B). The top layer is a Finite Mixture Model (FMM) (Equihua 1988), in which each *component* of the mixture represents a cluster of genomic segments that share temporal epigenomic patterns. Without considering the time factor, each component (cluster) degenerates into a set of

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

genomic segments sharing an epigenomic state. The bottom layer models the epigenomic data in each cluster. Each cluster is modeled as a HMM (Durbin et al. 1998) that represents the temporal changes of epigenetic modification intensities. The hidden states are binary activity states (inactive and active), which are allowed to change with respect to time. For example, if an enhancer changes from an active enhancer into an inactive enhancer during differentiation, the hidden states for this enhancer would change from 0 (inactive) to 1 (active). In a differentiation process, State 0 can be interpreted as the initial state before differentiation (Undifferentiated, U), whereas State 1 can be regarded as the other state in differentiated cells (Differentiated, D). The observed data are the epigenetic modification intensities for each genomic segment at every time point (Figure 1B). The sequencing reads from ChIP-seq experiments for each epigenetic modification on a genomic segment are modeled with a Poisson distribution, with the Poisson parameter reflecting the cluster and time-dependent epigenomic state. Thus, GATE has been completely specified as a generative probabilistic model (see Methods). In short, GATE is a finite mixture of Hidden Markov Models (HMM). Distinguishing itself from a previous method that predicts chromatin states based on static epigenetic data (Ernst and Kellis 2010), GATE models epigenomic data as a dynamic continuum and directly utilizes temporal information to annotate epigenomic states. This feature facilitates the discovery of cis-regulatory sequences and the annotation of their regulatory functions (Figure S6).

We implemented an Expectation-Maximization (EM) method (Dempster et al. 1977) to estimate model parameters from data (Figure S1, Supplementary Methods). At every maximization step, we embedded a Baum-Welch algorithm (Durbin et al. 1998) to estimate the HMM parameters. We provide GATE as a fully documented program at: http://biocomp.bioen.uiuc.edu/GATE. GATE was tested with simulated cell differentiation processes (Supplementary data 1, Tables S2-S4).

Epigenomic landscape during the differentiation of ES cells to mesendoderm

Mesendoderm is the diverging point of definitive endoderm and mesoderm (Yasunaga et al. 2005), which represents an important cell lineage besides the neural lineage (ectoderm) during the early stages of ES cell differentiation. We differentiated mouse ES cells to mesendoderm using Activin and a previously described culture medium (Yasunaga et al. 2005). On the sixth day of differentiation, almost all cells exhibited typical mesendoderm

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

morphology and expressed mesendoderm protein Goosecoid (GSC) (Blum et al. 1992) and endoderm protein SOX17 (Kanai-Azuma et al. 2002) (Figure S2A). Pluripotency genes *Pou5f1*(a.k.a. *Oct4*), *Sox2*, and *Nanog* were down-regulated, whereas endoderm and mesoderm genes *Gsc*, *Chordin*, *Foxa2*, *Sox17*, *Lim1*, and *Hnf4* were up-regulated (Figure S2B).

We measured a total of 9 epigenomic marks at 3 time points (Day 0, 4, and 6) during the differentiation process. These marks included 7 histone modifications or variants (H3K4me1/2/3, H3K27ac, H3K27me3, H3K36me3, and H2A.Z), which were assayed by chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Xiao et al. 2012). We supplemented the histone data with 2 types of DNA modifications, including 5-hmC by chemical labeling and pull-down followed by sequencing (5-hmC-seq) (Song et al. 2011) and 5-mC by both methylated DNA immunoprecipitation followed by sequencing (MeDIP-seq) and DNA digestion by methyl-sensitive restriction enzymes followed by sequencing (MRE-seq) (Maunakea et al. 2010). The 5-hmC pull-down specifically used the chemical property of the hydroxyl- group and thus was efficient to distinguish 5-hmC from 5-mC (Song et al. 2011). MeDIP-seq was representative of 5-mC, and MRE-seq was representative of unmethylated CpGs (^{u}CpG) (Maunakea et al. 2010).

To analyze the transcriptome, we sequenced ncRNAs using the Illumina Small RNA Sample Preparation procedure followed by sequencing (Illumina 2010b) and mRNAs using RNA-seq (Mortazavi et al. 2008) at the same 3 time points. Taken together, 36 sequencing datasets comprised of 1.94 billion 75 nt or 100 nt uniquely alignable sequencing reads were generated.

These data allowed us to estimate that 11.5% of the mouse genome is associated with at least one type of epigenetic modification in undifferentiated ES cells. Nearly half of these regions (5.60% of the genome) exhibited significant changes in at least one epigenetic modification during differentiation. About 1.92% of the genome was transcribed into mRNAs in ES cells, and 0.43% of the genome exhibited change of mRNA expression levels during differentiation.

Spatiotemporal clustering of epigenomic states

GATE clusters genomic segments based on both *spatial* distributions of epigenomic modifications and *temporal* changes of these modifications. Applying GATE to the “ES

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

cell to mesendoderm" differentiation dataset, we initially obtained 55 clusters, consistent with the previously estimated number of chromatin states (Ernst and Kellis 2010). These clusters formed 14 larger groups (Figure S3). Twelve of the 14 groups showed epigenomic characteristics that are typical to gene bodies, promoters, and enhancers. For example, Groups 2, 3, 6, and 9 shared enhancer characteristics including low H3K36me3, high H3K4me1 in either undifferentiated (U, Figure 2A) or differentiated states (D), and either high H3K27me3 or high H3K27ac (vertical green bar, Figure 2). Groups 1b, 5, 8, and 12 shared promoter characteristics including high H3K4me3 and low H3K36me3 (red bar). Groups 1a, 4, 10, and 11 shared high levels of H3K36me3, which was associated with gene bodies (Kolasinska-Zwierz et al. 2009; Schwartz et al. 2009) (blue bar, Figure 2).

By assigning clusters as promoter, enhancer, gene body, and repeat clusters (red, green, blue, and gray bars, Figure 2), we turned the unsupervised spatiotemporal epigenomic clusters into predictions of different genomic features. To check these predictions, we compared the locations of the genomic segments in every cluster to their nearest genes. Indeed, the relative locations of genomic segments in each cluster corroborated the unsupervised predictions (Figure S4). Chromosome 11 was randomly chosen for quantifying the prediction accuracies. The sensitivities for detecting promoters increased from 0 to 60% when the false positive rate (1-specificity) increased from 0 to 0.25% (Figure 3B). Similar tradeoffs between sensitivity and specificity were found for enhancer predictions (Figure 3C). These quantities reinforced the visual impression (Figure S4) that spatiotemporal clusters correlate with different genomic features. Changing the input size of genomic segments from 200 nt to 100 nt did not change any qualitative characteristics of the clustering results (Figure S5).

Spatiotemporal epigenomic clusters are predictive of transposons, bidirectional promoters, miRNA promoters, and piRNAs

The spatiotemporal epigenomic clusters did not only predict usual genomic features such as enhancers, promoters, and gene bodies, unexpectedly, these clusters were also capable of predicting genomic features including repeats, bidirectional promoters (Lin et al. 2007; Hakkinen et al. 2011), microRNA (miRNA) promoters (Marson et al. 2008), and PIWI RNAs (piRNA) (Thomson and Lin 2009).

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Two groups (Groups 13 and 14) exhibited unfamiliar epigenomic characteristics, including high H3K36me3 in parallel with high H3K27me3. These clusters did not locate in gene bodies or promoters (Figure S4). They contained the highest proportions of repeats among all 55 clusters (p-value = 5.45×10^{-37}, Fisher's exact test, "Repeats" column, Figure 2B). ncRNA expression showed that Groups 13 and 14 corresponded to transcribed and silenced repeat sequences (ncRNA, Figure 2B). These data suggest a distinct spatiotemporal epigenomic signature for repeats and transposons.

Bidirectional promoters were strongly enriched in Group 8 (Clusters 19, 12, 48) (p-value = 3.9×10^{-290}, Figure 3A). A simple but powerful classifier for identifying bidirectional promoters can be built based on the clusters. Based on whether a genomic segment belongs to Cluster 19, one can reach 80% sensitivity with a specificity of 99.6% (Figure 3D). Besides bidirectional promoters, miRNA promoters were also enriched in specific clusters, including Cluster 28 (p-value = 6.96×10^{-28}, Fisher's exact test) and Cluster 52 (p-value = 1.51×10^{-11}, Fisher's exact test), (Figure S7A). Thus, through unsupervised clustering, GATE revealed distinct spatiotemporal epigenomic patterns in several specific types of promoters.

PiRNAs and PIWI proteins were discovered in germ cells (Thomson and Lin 2009) and were thought to be silenced in ES cells. Unexpectedly, piRNAs were specifically enriched in Cluster 11 (p-value = 3.08×10^{-16}, Fisher's exact test) and Cluster 28 (p-value = 2.05×10^{-6}, Fisher's exact test) (Figure S7B). As a control, in Cluster 11 where piRNA was strongly enriched (9.53-fold more enriched than expected), miRNAs were depleted (0.89-fold less than expected). Thus piRNAs and miRNAs had different temporal epigenomic characteristics. These data suggest that even though piRNAs were produced and functional in germ cells, specific epigenomic patterns were formed on piRNA genes much earlier than germ cell development. PiRNA genes may be epigenetically prepared for activation in ES cells.

The distinct epigenomic characteristics in ES cells opened the possibility that a subset of piRNAs is produced in ES cells. Indeed, a cluster of piRNA genes (piRNA cluster) on Chromosome 5 was clearly expressed (Figure S7C). Moreover, the expression of this piRNA cluster was specifically induced in undifferentiated ES cells (Days 0, 4, 6, Figure S7C). As a control, Sgsm1, the neighboring gene to this piRNA cluster, showed a slight increase in expression during differentiation (Figure S7C). Even more strikingly, *Piwil2* (a.k.a. *Mili*), a mouse orthologue of the drosophila PIWI gene, is expressed in ES cells, and its expression decreased below a detectable level 4 days after differentiation (Figure

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

S7D). The consistent inductions of piRNAs and the Piwil2 gene in undifferentiated ES cells further entertained the hypothesis that some piRNAs were produced not only in germ cells.

Combinatorial temporal changes of epigenomic modifications

There were large differences among the nine epigenomic modifications regarding their temporal changes. The most stable modification was H3K36me3, which showed little change in all groups. The other eight modifications showed cluster-specific temporal changes (Figure 2).

Gene body groups showed three types of combinatorial changes. The first type was represented by coordinated decreases of H3K27ac, H3K4me1/2, H2A.Z, and 5-hmC, and an increase in CpG methylation (red circles in Group 4, Figure 2). These coordinated epigenomic changes did not clearly influence mRNA expression (mRNA, Figure 2B) but were associated with strong up-regulation of ncRNAs (ncRNA, Figure 2B). The second combination was decreases of H3K27ac and H3K4me2/3 (red circles, Group 11), which were associated with down-regulation of both mRNAs and ncRNAs (orange circles). The third combination was coordinated increases of H3K4me1/2 and 5-hmC, which corresponded to up-regulation of ncRNAs (Groups 1a and 10).

All promoter groups shared increases of H3K4me1/2, H2A.Z, 5-hmC, and 5-mC together with a decrease in CpG methylation. Interestingly, this recurrent combinatorial pattern itself was not sufficient to induce gene expression changes. However, when this pattern was further combined with a decrease in H3K4me3, it was associated with down-regulation of ncRNAs (Group 1b, Figure 2).

Enhancer groups showed four types of combinatorial changes. The first type was an increase in CpG methylation, which correlated with down-regulation of ncRNAs (red and orange circles, Group 6). The second combination was an increase in H3K27ac and decreases in H3K27me3 and 5-mC, which correlated with up-regulation of mRNAs and ncRNAs. The third type was coordinated decreases of H3K27ac, H3K4me1/2, H2A.Z, and 5-hmC together with an increase in 5-mC, which correlated with down-regulation of mRNAs and nRNAs (Group 9). The last type was coordinated decreases of H3K27ac and H3K4me3 together with a 5-mC increase, which correlated with down-regulation of mRNAs and ncRNAs (Group 7, Figure 2).

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Both transcribed and silenced repeats showed coordinated decreases in H3K27me3, 5-hmC, and 5-mC (especially non-CpG methylation), which correlated with up-regulation of transcription (ncRNA, Figure 2).

Recurrent themes of epigenomic and transcriptome changes

Recurrent themes appeared in the majority of the spatiotemporal clusters. These recurring patterns may represent basic properties of temporal gene regulation by the epigenome.

First, combinatorial epigenomic changes are prevalent. In every case except one, we observed combinatorial changes of 3 to 7 epigenomic modifications. Furthermore, different genomic features (promoters, enhancers, genes, and repeats) have different combinations of temporal changes.

Second, combinatorial patterns of epigenomic changes are predictive of gene expression changes (temporal axis, Figure 1A). While gene expression data were not used in clustering epigenomic data, the epigenomic clusters clearly distinguish gene groups with different temporal expression changes (Figure 2).

Third, almost all combinatorial epigenomic changes correspond with changes in ncRNA expression levels. The direction of ncRNA expression changes was not associated with the changes in any single epigenomic modification but was strongly predictable by combinatorial changes (Figure 6).

Fourth, epigenomic changes in enhancers instead of promoters are indicative of mRNA expression changes. This reconciles previous observations that temporal epigenomic changes were poorly correlated with gene expression changes during cell differentiation (Wu et al. 2011b) by reproducing such results in promoter regions with more epigenomic modifications; but also it points out the importance of epigenomic changes in enhancers. Consistently, human ChIP-chip analysis showed enhancer associated modifications including H3K27ac and H3K4me1 had greater dynamic changes than other modifications during ES cell differentiation (Hawkins et al. 2011).

Fifth, all assayed modifications except H3K36me3 have robust temporal changes in multiple genomic features (promoters, enhancers, gene bodies, and repeats). The robust and recurrent temporal changes appeared not only in genomic regions where the epigenomic modifications were abundant, but also in genomic regions where the

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

modification levels were low. Previously, H3K4me1 was associated with enhancers and H3K4me3 and H2A.Z were associated with promoters due to their abundance in these regions. However, H3K4me1 showed reliable changes not only in enhancers, but also in promoters and gene bodies where its modification level was low. Similarly, H3K4me3 (Group 7) and H2A.Z (Group 9) showed reliable changes in enhancers. These data may suggest regulatory functions of epigenomic modifications in previously ignored genomic regions.

Temporal changes of H3K4me2, non-CpG methylation, and H2A.Z are predictive of DNA hydroxymethylation

It remains unknown what guides TET enzymes to specific parts of the genome to convert 5-mC to 5-hmC. To explore the upstream signals that might specify where in the genome 5-mC should be converted to 5-hmC, we asked if there were any epigenomic modifications that correlate with 5-hmC in terms of temporal changes. Across all 55 epigenomic clusters (Figure 2), the temporal changes of 5-hmC were on average most correlated with H3K4me2, unmethylated CpG (^{u}CpG, measured by MRE-seq), H3K4me1, and H2A.Z (Figure S8). Next, we checked the temporal correlations between 5-hmC and every other assayed epigenomic mark on every genomic segment (200 nt window). On 78.1% of the genomic segments the temporal correlations between 5-hmC and H3K4me2 were larger than 0.8 (p-value$<10^{-300}$, Figure S9B). Strong temporal correlations between 5-hmC and ^{u}CpG, 5-hmC and H4K4me1, and 5-hmC and H2A.Z were also observed (Figure S9E, S9F). In contrast, H3K4me3 and H3K36me3 did not show large temporal correlations with 5-hmC. Categorizing genomic segments by their clusters, promoter segments showed the strongest temporal correlations between 5-hmC and H3K4me2, ^{u}CpG, and H2A.Z (figure S9). These data indicate strong associations between the di- and mono- methylation of H3K4 and hydroxylmethylation of nearby cytosines. The exchange of histone variants H2A and H2A.Z may also associate with 5-hmC synthesis.

The temporal changes of unmethylated CpG (^{u}CpG) and 5-hmC were strongly correlated (Figure S9K). Two scenarios can fit this data. First, the ^{u}CpG was generated by a cytosine demethylation preprocess that involves the conversion of 5-mC to 5-hmC (Bhutani et al. 2011). In other words, the *same* cytosine is converted from 5-mC into 5-hmC and then into C (same cytosine hypothesis). Alternatively, ^{u}CpG signals TET enzymes to the genomic neighborhood to convert neighboring 5-mC into 5-hmC (neighbor hypothesis).

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

If the same cytosine hypothesis is true, we would predict that 5-mC and 5-hmC are anti-correlated during ES cell differentiation. However, temporal changes of 5-mC did not anti-correlate with 5-hmC changes (Figure S9M). More genome segments had the same direction of 5-mC changes and 5-hmC changes than expected at random (Dashed line, Figure S9M, p-value<10^{-200}). These data are inconsistent with the same cytosine hypothesis. Conversely, TET1 contains a Znf_CXXC domain that interacts with ^{u}CpG, which is in line with the neighbor hypothesis.

We then explored the roles of non-CpG methylation (^{m}CpH). ^{m}CpH was reported in oocytes without known functions (Tomizawa et al. 2011). These ^{m}CpHs were presumably due to high levels of *de novo* methylation enzymes Dnmt3a/b in oocytes (Lees-Murdock et al. 2005). The simultaneous increase of 5-mC and decrease of CpG methylation in the same promoters (Figure 6, S4) suggest *de novo* non-CpG methylation. This is consistent with the increased expression levels of DMNT3b during guided differentiation of ES cells towards mesendoderm cells (Figure S9N). Temporal changes of ^{m}CpH were strongly correlated with 5-hmC changes (Figure S9D), suggesting the genomic segments undergoing non-CpG methylation were also experiencing hydroxymethylation. Complementary to these data, 5-hmC level was most enriched in low CpG regions in ES cells (Yu et al. 2012). In summary, temporal changes of H3K4me1/2, ^{m}CpH, and H2A.Z are predictive of 5-hmC changes throughout the mouse genome. These temporal correlations do not provide any causal information, but they may help to prioritize some hypotheses for future biochemical analyses of 5-hmC pathways.

H2A.Z is predictive of gene expression in a context-specific manner

Epigenomic marks were thought to be predictive of gene expression levels in a context independent manner (Karlic et al. 2010). The main support for this idea was that the linear model using epigenomic marks to predict expression can be learned from CD4+ T-cells and then applied to other types of T-cells (Karlic et al. 2010). To see if this theory holds during ES cell differentiation, we correlated the model-learned epigenomic modification levels to gene expression levels (Figure S10). Consistent with previous thoughts, some epigenomic marks were correlated with gene expression in a context independent manner, in the sense that these correlations persist in different genomic locations (enhancers, promoters, etc.) and in different differentiation states. Such context independent epigenomic marks included H3K36me3, H3K27ac, and 5-mC. However, several epigenomic marks showed context-specific influence on mRNA expression.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Promoter H3K4me3 was correlated with mRNA expression, but enhancer H3K4me3 was not (Figure S10). Promoter H2A.Z was not correlated with mRNA expression in undifferentiated ES cells (Pearson correlation < 0), but promoter H2A.Z was strongly correlated with mRNA expression after differentiation (Pearson correlation = 0.83) (Figure S10 Panel H2A.Z). Thus, epigenomic marks, although associated with gene expression, can have different directions of association in different cellular contexts.

A unified model for 5-hmC's effects on gene expression

The effects of 5-hmC on gene expression remain unclear. The difficulty of clarifying 5-hmC's effects lies, at least in part, in the seemingly conflicting data. On the one hand, promoter 5-hmC levels were anti-correlated with gene expression levels in both undifferentiated and differentiated ES cells (Figure S10 Panel 5-hmC, Figure 4A, B). This may suggest 5-hmC as a repressive mark. On the other hand, enhancer 5-hmC had a weak positive correlation with gene expression (Figure S10 Panel 5-hmC, Figure 4C, D), which may indicate an activation role. Is there a simple model that can accommodate all these data and illustrate 5-hmC's effects on gene expression? We identified a model by looking at temporal changes of 5-hmC.

5-hmC were concentrated on gene bodies in ES cells and were shifted to promoters after differentiation (Figures 4B, Figure S11). These data are consistent with the recent discovery of lower 5-hmC in promoters than in intragenic regions in ES cells (Booth et al. 2012). However, the large increase of promoter 5-hmC and the decrease of intragenic 5-hmC during ES cell differentiation were not reported before. Enhancer concentration of 5-hmC was increased during differentiation (Figure 4D). The enrichment of 5-hmC in enhancers and promoters in differentiated cells provide further support to the hypothesized regulatory role of 5-hmC (Ficz et al. 2011; Pastor et al. 2011; Stroud et al. 2011; Wu et al. 2011a).

5-hmC showed either a substantial increase or a moderate increase, but no decrease in any enhancer clusters or promoter clusters (Figure 4B, D). In both promoters and enhancers, a substantial increase of 5-hmC (promoter Groups 5 and 12 in Figure 4A, enhancer Groups 2, 3, 6 in Figure 4C) inevitably resulted in induction of gene expression.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

When 5-hmC did not substantially increase, nearby genes would show decreasing expression. This was consistent with both promoters (Group 8) and enhancers (Group 9). Similar analyses on MeDIP-seq and MRE-seq data suggest that 5-mC did not confound the observed effect of 5-hmC on gene expression (Figure S12). Thus, a simple model can explain the relationships between 5-hmC and gene expression during ES cell differentiation. If 5-hmC increases above a critical value, gene expression increases; otherwise gene expression decreases. This model states an association between 5-hmC and gene expression changes, and it does not rule out additional epigenomic marks as confounding factors.

Epigenomic states correlate with transcription networks

The GATE model infers epigenomic states of every genomic segment as undifferentiated (U) or differentiated (D), indicating when a genomic segment may change its regulatory functions. Thus, the GATE model provides a genome-wide view of the locations of regulatory sequences as well as their time of activation. Such information may help to clarify the transcription network (Hawkins et al. 2011). To explore this potential, we did a case study for three mesendoderm genes: *Fgf8*, *Sox17*, and *Foxa2*. A set of Group 3 enhancers were found in the introns and 3' intergenic regions of *Fgf8* (yellow boxes, Figure 5A). These enhancers were predicted to shift from inactive to active epigenomic states (U $\rightarrow$ D). Transcription factor binding sites (TFBS) of GSC and IRF-1 appeared in these enhancers (Figure 5B,C). Both GSC and IRF-1 are key regulators of mesendoderm differentiation (Blum et al. 1992; Bruce et al. 2007). An isolated enhancer was identified ~50,000 nt upstream to the *Sox17* gene (Figure 5D, S12). Another enhancer was found ~7,000 nt upstream to the *Foxa2* gene (Figure 5E). Using epigenomic data, GATE suggested that both the *Sox17* enhancer and the *Foxa2* enhancers shifted from inactive to active states (Figure S13), which was in line with the increased expression of these genes (FPKM, Figure 5D, E). A strong FOXA2 TFBS appeared in the *Sox17* enhancer, and a strong SOX17 TFBS appeared in a *Foxa2* enhancer. These data suggest a positive feedback loop between *Sox17* and *Foxa2*. Coincidently, a peak in a FOXA2 ChIP-seq experiment in mouse liver (GEO accession number: GSM427089 (MacIsaac et al. 2010)) co-localized with the GATE predicted *Sox17* enhancer (yellow bar, Figure S14A). This is the strongest peak (p-value < 10^{-6}) in the 70,000 nt sequence neighborhood of the *Sox17* gene. Moreover, the predicted FOXA2 TFBS appeared at the center of this peak (insert,

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Figure 14A). Reversely, when *Sox17* expression was induced in mouse ES cells, a strong peak (fold change = 8.1, p-value < 2.5×10^{-7}) of SOX17 ChIP-chip (GEO accession number: GSM470844 (Niakan et al. 2010)) appeared ~7,000 nt upstream to the *Foxa2* gene, co-localizing with the GATE predicted *Foxa2* enhancer (Figure S14B). Moreover, the predicted SOX17 TFBS located precisely at the center of this peak (insert, Figure 1B). These ChIP-seq/chip data reinforced the GATE predicted feedback loop. This feedback loop can stabilize the activation of two master transcription factors, and thus may be essential for mesendoderm differentiation.

Discussion

Previous computational methods for analyzing epigenomes relied primarily on spatial information of epigenomic marks. For example, an HMM model was developed to annotate genomic sequences by co-localization of multiple epigenomic marks (Ernst and Kellis 2010). GATE connects with the Ernst-Kellis model in that with only one time point, GATE degenerates into a zero-order HMM. Unlike the Ernst-Kellis model, though, GATE did not rely on prior information about the arrangements of genomic features. The unsupervised nature of GATE makes it capable of predicting the genomic features that were not included in a training process. In the ES cell differentiation process, GATE predicted bidirectional promoters, miRNA promoters, and piRNA genes with high accuracies.

Temporal information is as important as spatial information in studying epigenomic functions. A case in point is that although TET was known to interact with trithorax homolog MLL (Tahiliani et al. 2009), the MLL targets H3K4me1/2 were not pursued as a major clue for guiding TETs to specific genomic regions. This was probably due to the lack of a very strong spatial correlation between H3K4me1/2 and 5-hmC in any studied cell types. Indeed, 5-hmC was enriched not only in enhancers where H3K4me1/2 levels were high, but also in promoters (Yu et al. 2012), CTCF binding sites (Yu et al. 2012), and gene bodies (Booth et al. 2012) where H3K4me1/2 levels were not necessarily high. However, temporal correlations between H3K4me1/2 and 5-hmC were particularly strong, in that more than 85% of the genomic sequences had the same direction of changes of H3K4me2 and 5-hmC (Figure S9B, E). This result prioritizes H3K4me1/2 and H3K4 methyltransferases as a candidate upstream signal for guiding TETs to specific parts of the genome.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

People questioned ^{u}CpG's capability to attract TET1 (Frauer et al. 2011), despite that TET1 contains a zinc finger CXXC domain that can bind ^{u}CpG (Tahiliani et al. 2009). Indeed, no genome-wide mapping has showed strong overlaps of 5-hmC and ^{u}CpG. Furthermore, the information content of CpG is small, making it hard to believe that such a weak sequence signal can confer specificity to guide TET1. In this study, we reported striking temporal correlations of 5-hmC and ^{u}CpG throughout the genome, highlighting the necessity of analyzing epigenome dynamics and providing genome-wide data to support the role of ^{u}CpG in guiding TET1. TET1's interacting partner MLL contains a zinc finger CXXC domain as well. Theoretically, the MLL-TET1-^{u}CpGs three way interaction can be a lot more stable than a two way interaction of a protein and its DNA recognition site (He et al. 2009). This MLL-TET1-^{u}CpGs interaction is reinforced by MLL's roles to methylate H3K4 and interact with methylated H3K4. These analyses provide a model that ^{u}CpGs guide TET1 to specific genomic locations by initiating self-reinforcing ^{u}CpGs-MLL-H3K4me1/2-TET1 interactions.

It remains controversial whether 5-hmC predominantly exists in the CpG context. Stand bias analysis suggested presence of 5-hmC on CpH in ES cells (Ficz et al. 2011). However, this result was not supported by single-base resolution mapping of 5-hmC (Yu et al. 2012). We observed positive temporal correlations of ^{m}CpH and 5-hmC in multiple genomic regions. Two scenarios fit this observation. First, 5-hmC existed on CpH; alternatively, the temporal changes of ^{m}CpH were associated with 5-hmC changes in nearby CpGs. In promoters, ^{u}CpG levels measured by MRE-seq often increased as 5-hmC levels increased, which suggests that at least a subset of the newly converted 5-hmC in promoters were on CpHs. These results suggest perhaps examining differentiated ES cells may resolve the controversy of the presence of 5-hmC on CpHs.

Modeling and analyzing the temporal data facilitates the discovery of context-specific functions of epigenomic marks. In yeast, H2A.Z was associated with both active and inactive genes (Raisner et al. 2005). In this study, H2A.Z was clearly associated with active genes in mesendoderm cells but not in undifferentiated ES cells. This shows the importance of considering the cellular contexts when inferring the regulation functions of an epigenomic mark.

Most of the epigenomic marks (all assayed except H3K36me3) showed robust temporal changes in multiple genomic features, including promoters, enhancers, gene bodies, and repeats. This recurring theme can have large implications for studying gene regulation. In the canonical view, certain modifications are indicative of certain genomic features; for

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

example, H3K27ac and H3K4me1 are enhancer marks and H3K4me1 is a promoter mark. This canonical view was built on the observation that these modifications were a lot more abundant in certain genomic features than others. This view made it tempting to ignore the regulatory roles of modification in places where it is not abundant. However, the robust temporal changes of many modifications in their non-canonical (low-abundance) regions, such as H3K27ac in gene bodies, H3K4me1 in gene bodies and promoters, and H3K4me3 in gene bodies and enhancers, indicate that they can play regulatory roles in more genomic features than in the canonical view. Future experiments are needed to test this hypothesis. Consistent with this new view, the temporal epigenomic changes in low-abundance regions were sometimes correlated with mRNA or ncRNA expression changes (Figures 2 and 6).

Another dilemma in the epigenomic field is as follows. On the one hand, epigenomic changes are essential to organismal development, supported by the fact that different cell types exhibit clearly different epigenomic patterns (Hawkins et al. 2010; Ernst et al. 2011). Thus, the epigenome is expected to regulate gene expression during development and differentiation (Hawkins et al. 2011). However, on the other hand, temporal epigenomic changes during cell differentiation were reported to not correlate with gene expression changes (Wu et al. 2011b). Our method and data allow us to investigate this dilemma from a new perspective.

From the methodological perspective, GATE provides two advantages. First, the unsupervised clustering summarizes the combinatorial changes of multiple epigenomic marks. Previously, one had to compare gene expression changes with every epigenomic mark one-by-one, resulting in inconclusive or even conflicting results. This was because the association between gene expression and an epigenomic mark may be confounded by other epigenomic marks. GATE enables us to correlate gene expression changes with the combinatorial changes of multiple epigenomic marks. Second, GATE makes unsynchronized changes in different parts of the genome comparable. This enables us to effectively pull information together from different genomic segments with similar but unsynchronized temporal changes.

Consistent with the previous erythroid differentiation study (Wu et al. 2011b), temporal epigenomic changes did not correlate with gene expression changes in several GATE clusters. These clusters were all in promoters (Groups 5, 8, 12, Figure 2). However, temporal epigenomic changes in enhancer-associated clusters were clearly correlated with gene expression changes. In particular, changes in DNA methylation alone were

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

associated with ncRNA expression change (Group 6); changes that involve different combinations of DNA and histone modifications were associated with both mRNA and ncRNA changes in different directions (Groups 3, 7, 9, Figure 2); changes in modifications on repeats were predictive of repeat expression (Groups 13, 14). Including non-uniquely-mapped reads into the analysis may impact the results on repeat regions. These data suggest that the epigenome mediated gene regulation during cell differentiation, although clear in enhancers, may not be discernible in promoters, thus helping to resolve the hitherto mentioned dilemma.

Materials and Methods

The probabilistic model for analyzing temporal epigenomic data (GATE)

Symbols

Indices, $\boldsymbol{w}$: genomic segments; $\boldsymbol{t}$: time points; $\boldsymbol{m}$: epigenomic marks; $\boldsymbol{k}$: epigenomic clusters. **Observed data,** $\boldsymbol{W}$: the number of genomic segments; $\boldsymbol{M}$: the number of epigenomic marks; $\boldsymbol{T}$: the number of time points; $\boldsymbol{v}$: normalized sequence counts. Note that $\boldsymbol{v_{w,t,m}}$ is the normalized sequence count for epigenomic mark *m* in genomic segment *w* at time *t*. $\boldsymbol{O}$: all the observed data. **Hidden variables,** $\boldsymbol{C_w}$: cluster membership of the genomic segment *w*; $\boldsymbol{H}$: activity states, taking values 0 or 1. **Pre-computed parameters,** $\boldsymbol{K}$: the number of epigenomic clusters. **Model parameters,** $\boldsymbol{\pi_k}$: the proportion of genomic segments in cluster *k*; $\boldsymbol{b_{i,j}}$: transition probability from state *i* to state *j*; $\boldsymbol{\lambda}$: the Poisson parameter for emission probabilities; $\boldsymbol{\Lambda}$: all the model parameters.

The model

The GATE model is a hierarchical model with two layers (Figure 1B). The top layer is a Finite Mixture Model (FMM), in which each *component* of the mixture represents a cluster of genomic segments that share temporal epigenomic patterns. The bottom layer models the time-course epigenomic data in each cluster. Each cluster is modeled as a Hidden Markov Model (HMM) with the hidden states representing the changes of regulatory activities during a differentiation process, which emit the observed epigenomic data at each time point. Overall, GATE is a finite mixture of HMMs. **The top layer.** The top layer FMM models the cluster memberships of every genomic segment. The cluster membership of genomic segment $\boldsymbol{w}$ is modeled as a categorical distribution with probability $\pi = (\pi_1, \ldots, \pi_K)$:

$$C_w \sim Categorical(\pi), \text{ and } P(C_w) = \sum_{k=1}^{K} P(C_w = k)\pi_k \quad (1).$$

The HMM at the bottom layer. Given the cluster membership $\boldsymbol{C_w}$, the potential changes of regulatory activities for genomic segment *w* are modeled as a Markov chain. As a

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

hidden variable, $H_{w,t} \in (0, 1)$ represents the activity state of genomic segment *w* at time *t*.

The transition probability matrix (b^{C_w}) is written as:

$$b_{i,j}^{C_w} = P(H_{w,t+1} = j | H_{w,t} = i, C_w) \text{ , where } i, j \in (0, 1) \quad (2).$$

The conditional probability of H_w given C_w is:

$$P(H_w | b^{C_w}, C_w) = P(H_{w,1} | C_w) \prod_{t=1}^{T-1} P(H_{w,t+1} | H_{w,t}, C_w) = P(H_{w,1} | C_w) \prod_{t=1}^{T-1} b_{H_{w,t}, H_{w,t+1}}^{C_w} \quad (3).$$

Given the hidden variable, the observed sequence count for each epigenomic mark is modeled to follow a Poisson distribution (emission distribution). The Poisson parameter dependents on the cluster membership and the hidden regulatory state.

$$\nu_{w,t,m} \sim Poisson\left(\lambda^{C_w}{}_{H_{w,t},m}\right), \text{ where } 1 \le m \le M, 1 \le w \le W, 1 \le t \le T \quad (4).$$

Conditional on the cluster membership and the hidden variable, the different epigenomic marks are modeled as independent, and thus:

$$P(\nu_w | C_w, H_w; \lambda^{C_w}) = \prod_{t=1}^{T} P(\nu_{w,t} | C_w, H_{w,t}; \lambda^{C_w}) = \prod_{t=1}^{T} \prod_{m=1}^{M} P(\nu_{w,t,m} | C_w, H_{w,t}; \lambda^{C_w})$$

, where $\nu_w = (\nu_{w,1}, \nu_{w,2}, \ldots, \nu_{w,T})$ and $\nu_{w,t} = (\nu_{w,t,1}, \nu_{w,t,2}, \ldots, \nu_{w,t,M})$

(5).

Thus, a generative probabilistic model for all data has been fully specified.

Likelihood function

Under model assumptions, the likelihood function of observed data ($\boldsymbol{O}$) is

$$P(\boldsymbol{O} | b, \lambda, \pi) = \prod_{w=1}^{W} P(\nu_w | b, \lambda, \pi) = \prod_{w=1}^{W} \left(\sum_{C_w=1}^{K} P(\nu_w | C_w; b^{C_w}, \lambda^{C_w}) P(C_w) \right)$$
$$= \prod_{w=1}^{W} \left(\sum_{C_w=1}^{K} \left(P(C_w) \sum_{H_w} \left[P(\nu_w | C_w, H_w; \lambda^{C_w}) P\left(H_w | b^{C_w}, C_w\right) \right] \right) \right)$$

, where $b = (b^1, b^2, \ldots, b^K)$, $\lambda = (\lambda^1, \lambda^2, \ldots, \lambda^K)$ (6).

Parameter inference

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

The hidden variables of interest (C, H) were estimated by maximum likelihood estimation. We implemented a nested expectation-maximization (EM) algorithm in which the transition and emission parameters of the HMM were estimated by a Baum-Welch algorithm (Figure S1 and Supplementary Methods).

Fitting data to the GATE model

For simulation data, the cluster number (K) was estimated by the Bayesian Information Criteria (BIC):
$BIC = -2 * Q\left(\Lambda \middle| \Lambda^{(final)}\right) + (2 \times K - 1 + K \times M \times 2) \times ln(W)$, where $\Lambda = \{b, \lambda, \pi\}$ is the collection of all parameters.

For real data, we initial run the model with a relatively large cluster number (55), which was estimated from a previous study. The model-generated clusters were then merged into larger groups based on hierarchical clustering (Figure S3). We set the hidden state at the first time point as 0. We set $b_{1,0}^{C_w} = 1 - \ b_{1,1}^{C_w} = 0$, because it is unlikely to make two switches of regulatory states within this short differentiation time course.

Maintenance of ES cells

Undifferentiated mouse E14 ES cells were cultured under feeder-free conditions. ES cells were plated on gelatin-coated dishes with Dulbecco's modified Eagle medium (DMEM; GIBCO), supplemented with 15% heat-inactivated fetal bovine serum (FBS; GIBCO), 0.055 mM mercaptoethanol (2-ME; GIBCO), 2 mM L-glutamine, 0.1 mM MEM nonessential amino acid, 5,000 units/ml penicillin/streptomycin and 1,000 units/ml of LIF (Millipore ESG1107) at 37 ºC in 5 % CO_2 (Li et al. 2011).

Differentiation of ES cells into mesendoderm

Guided differentiation of ES cells was performed as described before (Yasunaga et al. 2005). Briefly, 2×10^5 cells were seeded on Collagen IV-coated 10-cm dishes (BD, 08-774-33) in serum free medium ESF-B (Itochu Corporation), supplemented with 0.1 % BSA, 50 µM 2-ME and 10 ng/ml Activin A (R&D Systems). The culture medium was changed every day.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Reverse transcription polymerase chain reaction (RT-PCR)

Total RNA was extracted from cells 0, 4, and 6 days after induction of differentiation with TRIzol® Reagent (Invitrogen, 15596-026) according to manufacturer's protocol. First-strand cDNA was prepared with M-MLV Synthesis System (New England Biolabs). Beta-actin was used as a control. Pluripotency and lineage-specific marker genes were assayed by Applied Biosystems 7900HT Fast Real-Time PCR System. Primer sequences are listed in Table S5. Gene expression levels were normalized to the expression level of β-actin. Fold changes of gene expression levels were calculated between Day 4 and Day 0, and between Day 6 and Day 0.

Immunostaining

A total of 10,000 - 20,000 cells were seeded on Collagen IV-coated 35-mm dishes (Ibidi 45074) in the medium used for guided differentiation (ESF-B supplemented with 0.1 % BSA, 50 μM 2-ME and 10 ng/ml Activin A). Cells were fixed in 4 % Paraformaldehyde. Primary antibodies against Goosecoid (Origene, TA500087) and Sox17 (Millipore 09-038) were mixed and applied to the fixed cells for 2 hours at 37 °C. The goat anti-mouse IgG antibody conjugated with Alexa 568 (Invitrogen A-11031) and the goat anti-rabbit IgG antibody conjugated with Alexa 488 (Invitrogen A-11034) were sequentially applied to the samples, each for 2 hours at 37 °C, as secondary antibodies. Cellular nuclei were stained by Hoechst 33342 (Invitrogen, H3570) for 15 minutes at room temperature. Images were obtained using LSM 700 microscope (Zeiss).

Chromatin immunoprecipitation

Cells were fixed with 1 % formaldehyde for 10 minutes at room temperature. Fixation was inactivated by addition of 125 mM glycine. Cells were scraped off of dishes and collected by centrifugation. Cross-linked chromatin-DNA complexes were isolated from the nuclei lysis buffer and then sonicated into fragments of a size range between 350 and 600 nt. Specific antibodies listed below were incubated with the solubilized DNA fragments at 4 °C overnight.

Antibodies used in this work are as follows: anti-histone H2A.Z antibody - ChIP Grade (Abcam, ab4174); anti-histone H3 (mono methyl K4) antibody - ChIP Grade (ab8895); anti-histone H3 (di methyl K4) antibody [Y47] - ChIP Grade (ab32356); anti-histone H3 (tri methyl K4) antibody - ChIP Grade (ab1012); anti-histone H3 (tri methyl K4) antibody - ChIP Grade (ab8580); anti-histone H3 (tri methyl K36) antibody - ChIP Grade (ab9050); anti-histone H3 (acetyl K27) antibody - ChIP Grade (ab4729); anti-H3K27me3 antibody (Millipore, 07-449); and a monoclonal antibody against 5-methylcytindine (Eurogentec, bi-mecy-0100).

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Antibody-chromatin complexes were captured by protein A/G agarose beads (Pierce, Rockford, IL), and eluted with 1 % SDS after extensive washing. The cross-link between DNA and chromatin proteins was reversed by incubation in 20 mM NaCl at 65 °C overnight. DNA was purified by QIAquick PCR Purification Kit (Qiagen 28106) and dissolved into 30 µl TE buffer per immunoprecipitation.

MRE-seq, MeDIP-seq, and 5-hmC-seq

MRE-seq and MeDIP-seq were performed as described before (Maunakea et al. 2010). Sample DNA was digested in parallel using HpaII, Hin6I, SsiI (Fermentas), and HpyCH4IV (New England Biolabs) before deep sequencing.

5-hmC-seq was performed as described before (Song et al. 2011). 5-hmC was chemically labeled as selected. Sequencing libraries were constructed using the same protocol as ChIP-seq (Illumina 2010a).

Small RNA sequencing

Total RNA was purified with TRIzol® Reagent (Invitrogen 15596-026) and used as an input to generate a small RNA library using True-seq small RNA kit (Illumina RS-200-0012). The RNA 3' adapters in this kit were specifically modified to target microRNAs and other small RNAs that have a 3' hydroxyl group resulting from enzymatic cleavage by Dicer or other RNA processing enzymes. Library products ranging from 145 nt to 160 nt were collected and amplified. The libraries were quantitated by qPCR, and subsequently sequenced on a HiSeq2000 sequencer using TruSeq SBS sequencing kit version 2 and analyzed with pipeline version 1.8 (Illumina).

Data access

All sequencing data are accessible from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under SuperSeries accession number GSE38596.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

References

Beer MA, Tavazoie S. 2004. Predicting gene expression from sequence. *Cell* **117**(2): 185-198.

Bernstein BE, Meissner A, Lander ES. 2007. The mammalian epigenome. *Cell* **128**(4): 669-681.

Bhutani N, Burns DM, Blau HM. 2011. DNA Demethylation Dynamics. *Cell* **146**(6): 866-872.

Blum M, Gaunt SJ, Derobertis EM. 1992. Gastrulation in the Mouse - the Role of the Homeobox Gene Goosecoid. *Mol Biol Cell* **3**: A98-A98.

Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, Balasubramanian S. 2012. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**(6083): 934-937.

Bruce SJ, Gardiner BB, Burke LJ, Gongora MM, Grimmond SM, Perkins AC. 2007. Dynamic transcription programs during ES cell differentiation towards mesoderm in serum versus serum-freeBMP4 culture. *BMC Genomics* **8**: 365.

Buchler NE, Gerland U, Hwa T. 2003. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A* **100**(9): 5136-5141.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**(50): 21931-21936.

Dempster AP, Laird NM, Rubin DB. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society* **39**(1): 1-38.

Durbin R, Eddy SR, Krogh A, Mitchison G. 1998. *Biological Sequence Analysis*. Cambridge University Press.

Equihua M. 1988. Analysis of Finite Mixture of Distributions - a Statistical Tool for Biological Classification Problems. *Comput Appl Biosci* **4**(4): 435-440.

Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**(8): 817-825.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**(7345): 43-49.

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**(7347): 398-U589.

Frauer C, Rottach A, Meilinger D, Bultmann S, Fellinger K, Hasenoder S, Wang M, Qin W, Soding J, Spada F et al. 2011. Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PloS one* **6**(2): e16627.

Hakkinen A, Healy S, Jacobs HT, Ribeiro AS. 2011. Genome wide study of NF-Y type CCAAT boxes in unidirectional and bidirectional promoters in human and mouse. *Journal of Theoretical Biology* **281**(1): 74-83.

Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S et al. 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**(5): 479-491.

Hawkins RD, Hon GC, Yang CH, Antosiewicz-Bourget JE, Lee LK, Ngo QM, Klugman S, Ching KA, Edsall LE, Ye Z et al. 2011. Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency. *Cell Res* **21**(10): 1393-1409.

He X, Chen CC, Hong F, Fang F, Sinha S, Ng HH, Zhong S. 2009. A biophysical model for analysis of transcription factor interaction and binding site arrangement from genome-wide binding data. *PloS one* **4**(12): e8155.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Illumina. 2010a. TruSeq(TM) DNA Sample Preparation Guide.

Illumina. 2010b. TruSeq(TM) Small RNA sample preparation guide.

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**(7310): 1129-U1151.

Kanai-Azuma M, Kanai Y, Gad JM, Tajima Y, Taya C, Kurohmaru M, Sanai Y, Yonekawa H, Yazaki K, Tam PPL et al. 2002. Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* **129**(10): 2367-2379.

Karlic R, Chung HR, Lasserre J, Vlahovicek K, Vingron M. 2010. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A* **107**(7): 2926-2931.

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* **41**(3): 376-381.

Lees-Murdock DJ, Shovlin TC, Gardiner T, De Felici M, Walsh CP. 2005. DNA methyltransferase expression in the mouse germ line during periods of de novo methylation. *Dev Dyn* **232**(4): 992-1002.

Li Y, Yokohama-Tamaki T, Tanaka TS. 2011. Short-term serum-free culture reveals that inhibition of Gsk3beta induces the tumor-like growth of mouse embryonic stem cells. *PLoS One* **6**(6): e21355.

Lin JM, Collins PJ, Trinklein ND, Fu YT, Xi HL, Myers RM, Weng ZP. 2007. Transcription factor binding and modified histones in human bidirectional promoters. *Genome Research* **17**(6): 818-827.

MacIsaac KD, Lo KA, Gordon W, Motola S, Mazor T, Fraenkel E. 2010. A quantitative model of transcriptional regulation reveals the influence of binding location on expression. *PLoS Comput Biol* **6**(4): e1000773.

Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J et al. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**(3): 521-533.

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong CB, Nielsen C, Zhao YJ et al. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**(7303): 253-U131.

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

Niakan KK, Ji H, Maehr R, Vokes SA, Rodolfa KT, Sherwood RI, Yamaki M, Dimos JT, Chen AE, Melton DA et al. 2010. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Dev* **24**(3): 312-326.

Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P et al. 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**(7347): 394-397.

Pearl J. 1985. Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning. In *7th Conference of the Cognitive Science Society*, p. 6, University of California, Los Angeles.

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

Raisner RM, Hartley PD, Meneghini MD, Bao MZ, Liu CL, Schreiber SL, Rando OJ, Madhani HD. 2005. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* **123**(2): 233-248.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Schwartz S, Meshorer E, Ast G. 2009. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* **16**(9): 990-U117.

Song CX, Szulwach KE, Fu Y, Dai Q, Yi CQ, Li XK, Li YJ, Chen CH, Zhang W, Jian X et al. 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* **29**(1): 68-+.

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

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L et al. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**(5929): 930-935.

Thomson T, Lin H. 2009. The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annu Rev Cell Dev Biol* **25**: 355-376.

Tomizawa S, Kobayashi H, Watanabe T, Andrews S, Hata K, Kelsey G, Sasaki H. 2011. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* **138**(5): 811-820.

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

Wu W, Cheng Y, Keller CA, Ernst J, Kumar SA, Mishra T, Morrissey C, Dorman CM, Chen KB, Drautz D et al. 2011b. Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. *Genome Res* **21**(10): 1659-1671.

Wyatt GR, Cohen SS. 1952. A new pyrimidine base from bacteriophage nucleic acids. *Nature* **170**(4338): 1072-1073.

Xiao S, Xie D, Cao X, Yu P, Xing X, Chen CC, Musselman M, Xie M, West FD, Lewin HA et al. 2012. Comparative epigenomic annotation of regulatory DNA. *Cell* **149**(6): 1381-1392.

Yasunaga M, Tada S, Torikai-Nishikawa S, Nakano Y, Okada M, Jakt LM, Nishikawa S, Chiba T, Era T. 2005. Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nat Biotechnol* **23**(12): 1542-1550.

Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B et al. 2012. Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell* **in press.**

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Figure legends

Figure 1. The Genomic Annotation based on Time-course Epigenomic data (GATE) model.
(A) Two types of correlations between the epigenome and gene expression. Spatial correlation (S) examines different genes in a fixed cell type, and temporal correlation (T) examines different differentiation stages or cell types for a fixed gene. Spatial correlation is often much more pronounced than temporal correlation.

(B-C) GATE models the genome as equal sized genomic segments, and each segment is associated with temporal epigenomic data. The model assumes that there are shared temporal epigenomic patterns among different genomic segments. GATE is a hierarchical model. The top layer is a Finite Mixture Model for clustering genomic segments. The bottom layer models the temporal changes within each cluster as a Hidden Markov Model. The hidden variables (circles) are a binary variables indicating the time of a change of regulatory activities. Emitted (vertical arrows) from the hidden variable are the intensities of each epigenomic mark.

Figure 2. Epigenomic clusters.
(A) Average intensities of each epigenomic mark (column) in each cluster (row). The model allows each sequence segment to have two activity states, denoted as undifferentiated ("U") and differentiated ("D"). The model assumes that each cluster has a shared mean intensity for each epigenomic mark at either activity state. These mean intensities are plotted in this matrix and are color coded. Clusters with similar intensities were merged into larger clusters (Groups. See Figure S3 for the merging procedure). Based on the epigenomic patterns, the groups were assigned with representative names, including promoters, enhancers, genes, and repeats (vertical color bars). Consistent intensity changes across multiple clusters in a group were highlighted (red circles).

(B) Fold enrichment of genomic and epigenomic features (column) in each cluster (row). Fold enrichment was calculated as the ratio between the average signal of a cluster to the average signal of all clusters. mRNA: transcription levels of nearest genes, derived from RNA-seq data. ncRNA: transcription levels of the genomic segment in each cluster, derived from ncRNA-seq data. CpG: CpG density. P300: P300 binding intensity. PolII: PolII binding intensity. Repeats: repeat density. Significant temporal changes were highlighted (orange circles).

Figure 3. Predicting genomic features.
(A) Distribution of bidirectional promoters in the 55 epigenomic clusters. Fold enrichment: the ratio between the percentage of bidirectional promoters in a cluster and the average percentage of all clusters. *: p-value $<10^{-60}$. **: p-value $<10^{-90}$.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

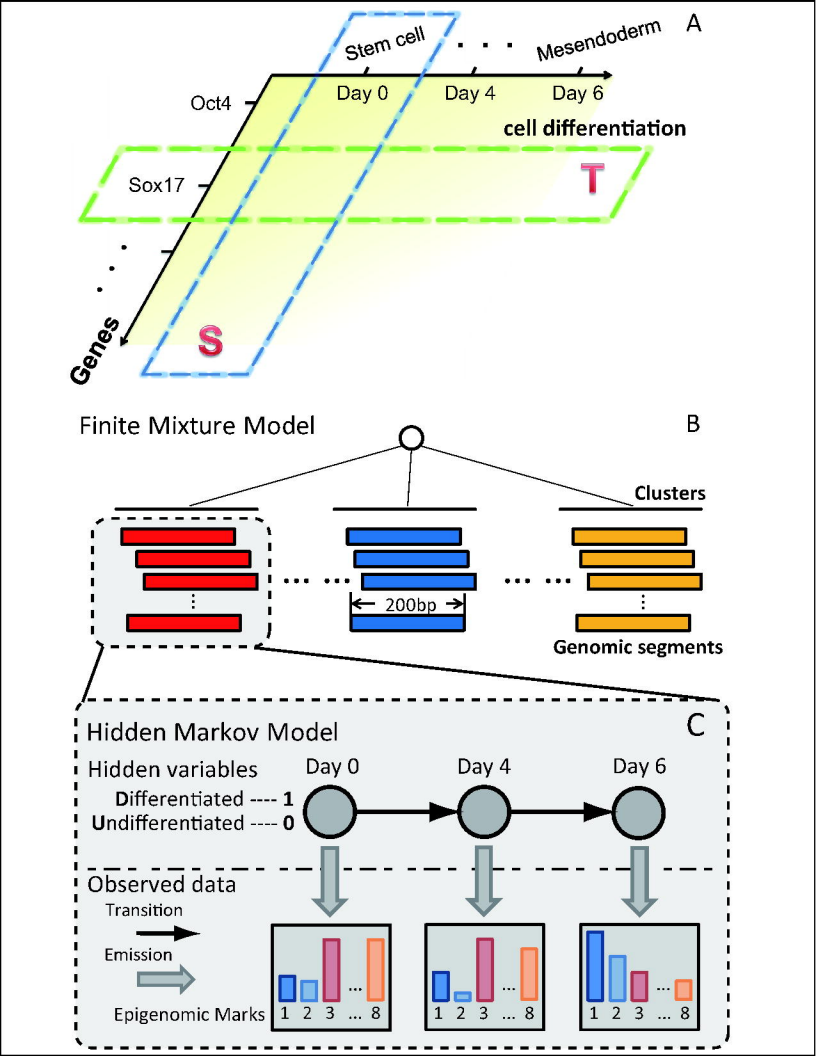
(B-D) Accuracies of predicting genomic features as measured by receiver operating characteristic (ROC) curves. AUC: area under the curve. Promoters were predicted by Groups 1b, 5, 8, and 12 (p-value < 2.2×10^{-16}, Wilcoxon test). Gene bodies were predicted by Groups 1a, 4, 10, and 11 (p-value < 2.2×10^{-16}). Bidirectional promoters were predicted by Cluster 19 (p-value < 2.2×10^{-16}). Inserts: details of the high specificity regions. (E) A predicted bidirectional promoter and a regular promoter. Along a fraction of Chromosome 11, each genomic segment is colored by the cluster it belonged to. The stretch of DNA belonging to Cluster 19 (yellow) corresponded to a bidirectional promoter.

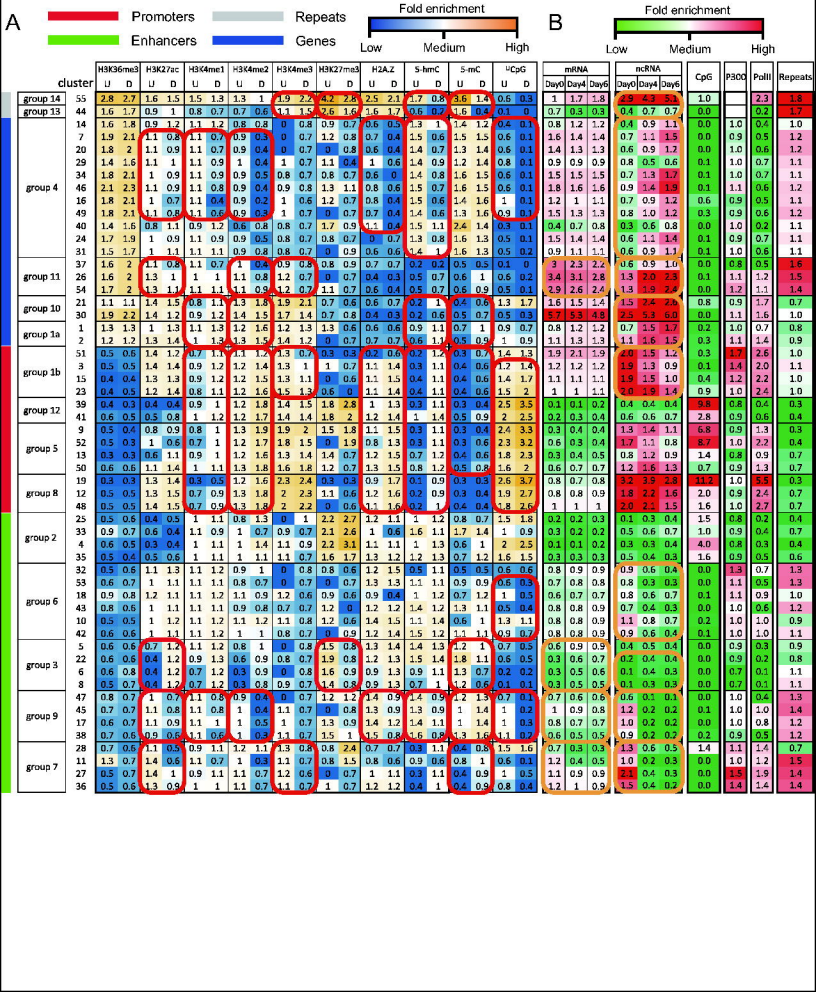
Figure 4. A unified model the effects of 5-hmC on gene expression.
(A-B) The average intensities of 5-hmC on predicted promoter groups (A) and enhancer groups (B) in three time points. 3000 base pairs (bp) of flanking regions centered on the centers of the genomic segments of each group were shown.

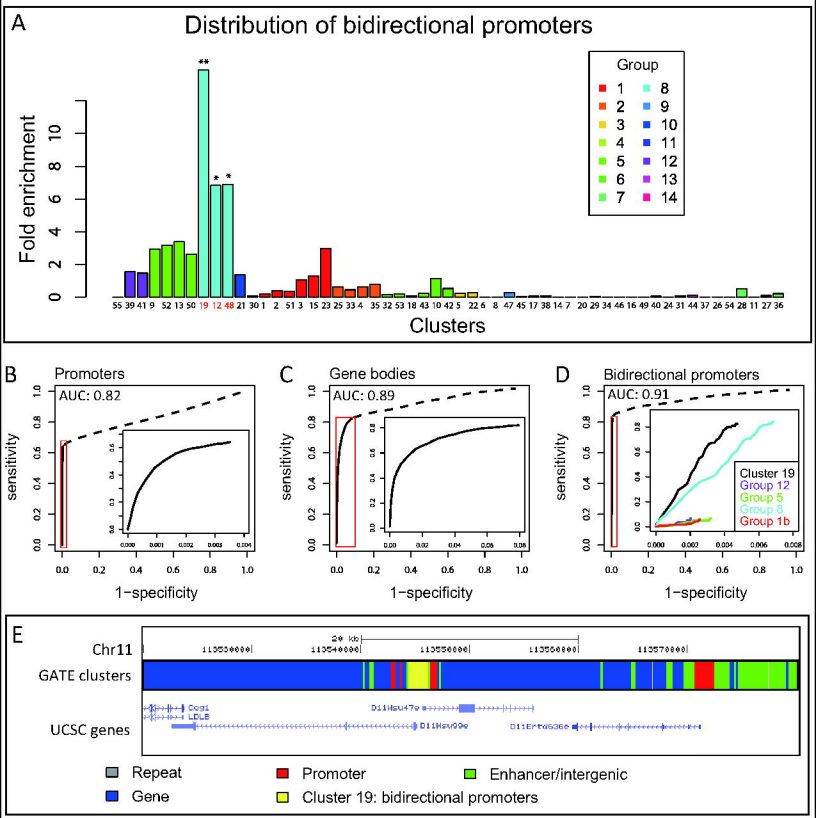
(C-D) The expression levels of nearby genes for predicted promoter groups (B) and enhancer groups (D). Mean and standard deviation (error bar) were shown for each time point. Whether 5-hmC increases above a critical threshold (horizontal line, A-B) correlates with the direction of expression changes. FPKM: fragments per kilobase of exon per million fragments mapped.

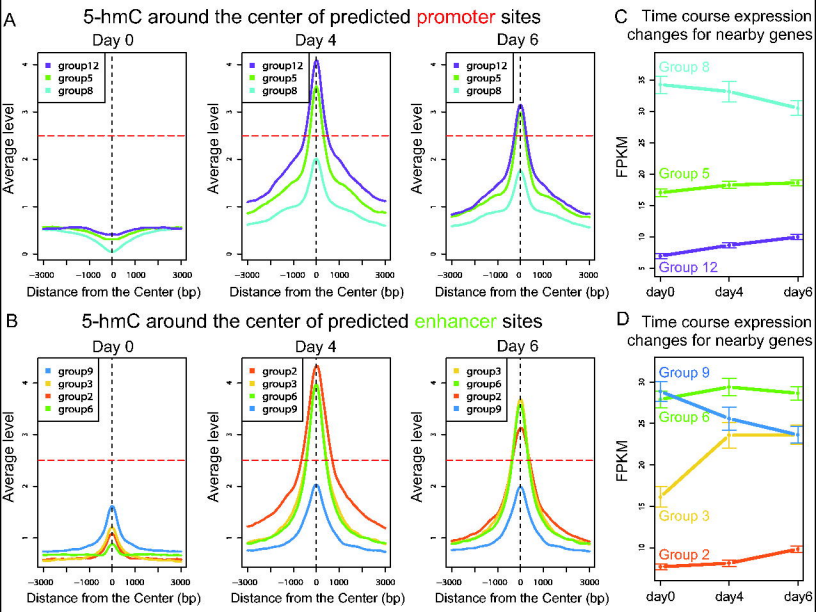
Figure 5. Predicted mesendoderm enhancers harbor transcription factor binding sites.
Epigenomic clusters near the *Fgf8* (A), *Sox17* (D), and *Foxa2* (E) genes. Genomic segments (colored bars) were marked by their cluster numbers on the left. Their variable widths indicate their activity states. A left-thin-right-fat bar indicates a change of the activity states. A strong GSC motif (B) and a strong IRF-1 motif (C) appeared in predicted enhancer segments in the 3' and the intron of the *Fgf8* gene. Both predicted enhancers showed changes of activities during the differentiation (left-thin-right-fat). A strong *Foxa2* motif appeared in a predicted enhancer 50,000bp upstream to the *Sox17* gene (D). In turn, a strong *Sox17* motif appeared in a predicted *Foxa2* enhancer. FPKM: fragments per kilobase of exon per million fragments mapped.

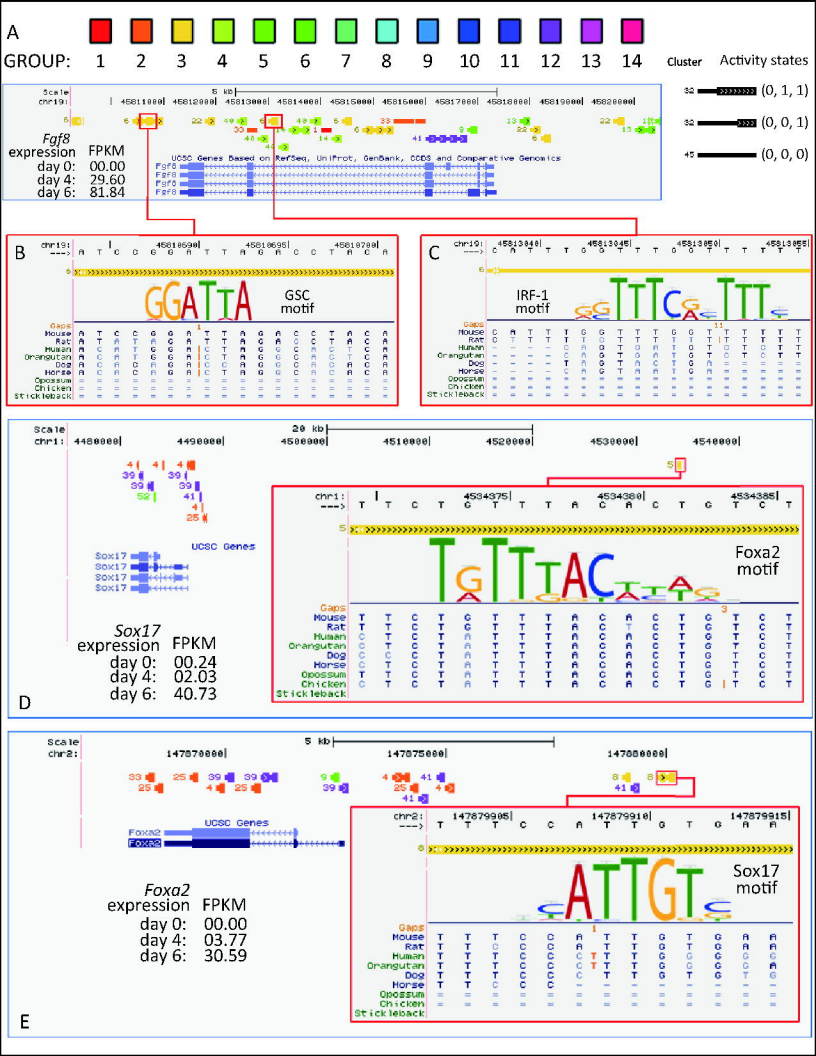
Figure 6. A model for combinatorial epigenomic changes and gene expression.
One to several combinatorial patterns (columns) were identified on each genomic feature. The model-learned intensity levels for each epigenomic mark (Input rows) in the undifferentiated state and the differentiated state were color-coded, and the directions of change were marked by arrows. The corresponding temporal changes of mRNA and ncRNA expression were color-coded in a green-to-red scale (Output rows).

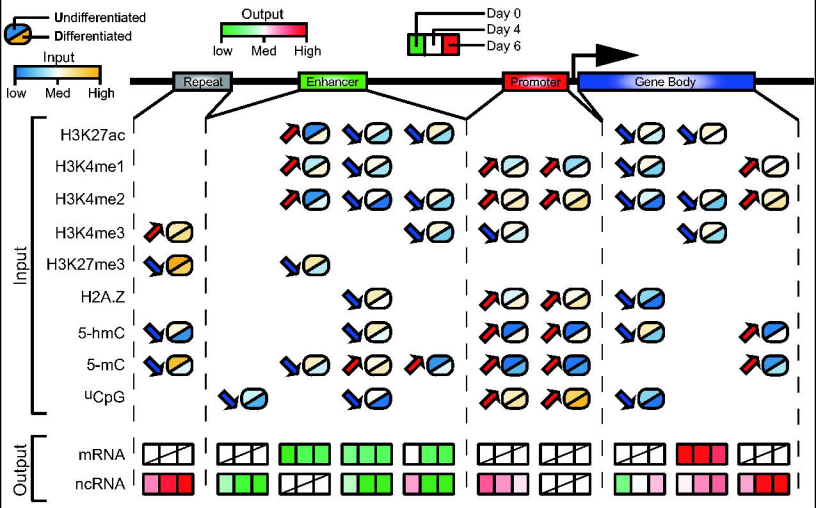

A
Stem cell
Mesendoderm
Day 0
Day 4
Day 6
Oct4
cell differentiation
Sox17
T
Genes
S
Finite Mixture Model
B
Clusters
200bp
Genomic segments
C
Hidden Markov Model
Hidden variables
Day 0
Day 4
Day 6
Differentiated ---- 1
Undifferentiated ---- 0
Observed data
Transition
Emission
Epigenomic Marks
1 2 3 ... 8
1 2 3 ... 8
1 2 3 ... 8


A
Promoters
Enhancers
Repeats
Genes
Fold enrichment
Low
Medium
High
B
Fold enrichment
Low
Medium
High
cluster
H3K36me3
H3K27ac
H3K4me1
H3K4me2
H3K4me3
H3K27me3
H2A.Z
5-hmC
5-mC
UCpG
mRNA
ncRNA
CpG
P300
PolII
Repeats
Day0
Day4
Day6
group 14
group 13
group 4
group 11
group 10
group 1a
group 1b
group 12
group 5
group 8
group 2
group 6
group 3
group 9
group 7


A
Distribution of bidirectional promoters
Fold enrichment
0 2 4 6 8 10
**
* *
Group
1 2 3 4 5 6 7
8 9 10 11 12 13 14
55 39 41 9 52 13 50 19 12 48 21 30 1 2 51 3 15 23 25 33 4 35 32 53 18 43 10 42 5 22 6 8 47 45 17 38 14 7 20 29 34 46 16 49 40 24 31 44 37 26 54 28 11 27 36
Clusters
B
Promoters
AUC: 0.82
sensitivity
1−specificity
C
Gene bodies
AUC: 0.89
sensitivity
1−specificity
D
Bidirectional promoters
AUC: 0.91
Cluster 19
Group 12
Group 5
Group 6
Group 1b
sensitivity
1−specificity
E
Chr11
20 kb
113550000 113560000 113570000
GATE clusters
Cog1
LDLB
UCSC genes
Repeat
Gene
Promoter
Cluster 19: bidirectional promoters
Enhancer/intergenic


A
5-hmC around the center of predicted promoter sites
Day 0
Day 4
Day 6
group12
group5
group8
Average level
Distance from the Center (bp)
C
Time course expression changes for nearby genes
Group 8
Group 5
Group 12
FPKM
day0
day4
day6
B
5-hmC around the center of predicted enhancer sites
group9
group3
group2
group6
D
Time course expression changes for nearby genes
Group 9
Group 6
Group 3
Group 2


A
GROUP: 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Cluster
Activity states
(0, 1, 1)
(0, 0, 1)
(0, 0, 0)
Fgf8
expression FPKM
day 0: 00.00
day 4: 29.60
day 6: 81.84
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
B
GSC motif
C
IRF-1 motif
D
Sox17
expression FPKM
day 0: 00.24
day 4: 02.03
day 6: 40.73
UCSC Genes
Foxa2 motif
E
Foxa2
expression FPKM
day 0: 00.00
day 4: 03.77
day 6: 30.59
UCSC Genes
Sox17 motif


Undifferentiated
Differentiated
Input
low Med High
Output
low Med High
Day 0
Day 4
Day 6
Repeat
Enhancer
Promoter
Gene Body
Input
H3K27ac
H3K4me1
H3K4me2
H3K4me3
H3K27me3
H2A.Z
5-hmC
5-mC
uCpG
Output
mRNA
ncRNA