Legends for supplementary figures

Figure S1. An expectation-maximization algorithm for inferring model parameters.

Figure S2. Differentiation of mouse ES cells into mesendoderm cells. (A) Immunofluorescence staining of cells at 0, 4, and 6 days of differentiation. DNA was stained by Hoechst in blue. GSC and SOX17 proteins were stained in red and green. At the 6^{th} day of differentiation, almost all cells express GSC and SOX17. (B) Comparison of mRNA expression of pluripotency and lineage-specific marker genes. Real time quantitative PCR analyses were carried out in three biological replicates at each time point. Fold change: the ratio of expression levels between Day 4 and Day 0 (yellow), and the ratio between Day 6 and Day 0 (green).

Figure S3. Dendrogram of epigenomic clusters. The hierarchical tree represents the relative distances between the clusters. Each leaf is an epigenomic cluster (cluster number shown below each leaf). Cutting the dendrogram at the height of 30, 14 groups were formed. Each group contained several clusters with similar spatiotemporal epigenomic patterns. Group 1a and Group 1b are the two main branches of Group 1. The input size of genomic segments for GATE clustering was 200 nt.

Figure S4. Genomic locations of epigenomic clusters. The relative location of every genomic segment with respect to the nearest gene was categorized (columns). These relative locations were summarized for each cluster (row). The relative abundance of each cluster (row) in each location category (column) is quantified by fold enrichment. The fold enrichments are shown in a green-red heatmap. The fold enrichment was defined as $\left(\frac{\#\text{ of segments of a cluster belonging to this category}}{\#\text{ of segments of this cluster}}\right) / \left(\frac{\#\text{ of segments belonging to this category}}{\#\text{ of segments of all clusters}}\right)$. The order of groups and clusters (rows) is the same as in Figure 2. Vertical color bars on the left indicate an unsupervised assignment of genomic features to the clusters.

Figure S5. The impact of the sizes of genomic segments to GATE clusters. The GATE clustering of epigenomic data were repeated with 100 nt as the new input size of genomic segments. The following qualitative characteristics of the new clustering results are the same as those of the 200 nt clustering results: (A) The 55 clusters can be hierarchically grouped into 14 groups; The clusters exhibit typical epigenomic patterns of promoters, enhancers, gene bodies, and repeats, and thus these clusters can be assigned as enhancer, promoter, gene, and repeat clusters; (B) The genomic locations of segments in each cluster corroborate with cluster assignments; The two "repeat" clusters have the largest fractions of repeat elements among all 55 clusters (Repeat column).

Figure S6. Activity states reflect temporal changes of enhancer activities. The activity states in the model allows different genomic segments in the same cluster to change their regulatory

activities at different time points. Taking Group 3 as an example, Group 3 was inferred to be an enhancer cluster. During the differentiation, the enhancer activity of each segment could change from inactive to active, or from active to inactive, between Day 0 and Day 4 (denoted as (0,1,1)) or between Day 4 and Day 6 (denoted as (0,0,1)). (0,0,0) denotes that the activity state does not change. (A) Expression levels of nearby genes correlate with model-inferred activity states. The induction of expression near (0,0,1) and (0,1,1) enhancers suggests that Group 3 contained mesendoderm enhancers. FPKM: fragments per kilobase of exon per million fragments mapped. (B) Temporal changes of three epigenomic marks on the (0,1,1) enhancers. Large changes happened between Day 0 and Day 4. (C) Enriched functional categories for the genes near the three types of enhancers. GO: gene ontology.

Figure S7. Distribution of miRNA promoters and piRNA genes in GATE clusters. Fold enrichment: the ratio between the percentage of miRNA promoters (A) or piRNA genes (B) in a cluster and the average percentage of all clusters. *: p-value $<10^{-5}$. **: p-value $<10^{-15}$. (C) An piRNA gene cluster on Chromosome 5. Cluster 11 genomic segments are enriched in this piRNA gene cluster. This piRNA gene cluster is expressed in ES cells (RNA-seq Day 0), but is not expressed after differentiation (Day 4, Day 6). The nearby gene, Sgsm1, shows a relatively constant expression. (D) Expression of PIWI protein genes *Piwil2* (Lane a), *Piwil2* (Lane b), *Piwil4* (Lanes c-f) in ES cells and during differentiation. FPKM: fragments per kilobase of exon per million fragments mapped.

Figure S8. Correlations of temporal changes. (A) Pairwise correlation between assayed epigenomic marks. These are correlations of temporal changes. The temporal changes of 5-hmC was correlated with those of H3K4me1/2, ^{u}CpG (MRE), and H2A.Z. (B) The calculation process of temporal correlations in Panel A. First, for each epigenomic mark, the difference of its intensities in the undifferentiated and the differentiated states was calculated in every cluster. This result was stored in a column vector. The correlation between two epigenomic marks was the Pearson correlation of the two column vectors representing the two epigenomic marks.

Figure S9. Temporal correlations between 5-hmC and other epigenomic marks. (A) Distribution of Pearson correlation coefficients between 5-hmC and the expression of nearby genes (the nearest gene within 100,000 bp for each segment). (B-L) Distribution of Pearson correlation coefficients between 5-hmC and other epigenomic marks. The distributions were separately calculated for genomic segments in promoter (red), enhancer (green), and gene body clusters (blue). A background distribution was derived by permuting the data from the three time points (dotted line). (M) The calculation procedure for the distributions in Panels A-L. First, on every genomic segment, a correlation was calculated between two marks using their intensities on three time points. Second, the correlations on all genomic segments were summarized into a histogram. (N) Expression levels of DNA methylation enzymes *Dmnt3a* (different transcript isoforms in Lanes a-b) and *Dmnt3b* (Lanes c-n). Several *Dmnt3b* transcripts showed increased expression during differentiation. FPKM: fragments per kilobase of exon per million fragments mapped.

Figure S10. Correlations between gene expression and epigenomic marks are context-dependent. (A) Correlations of each epigenomic mark with gene expression. U: undifferentiated state. D: differentiated state. (B) Calculation procedure. In each kind of clusters (promoter, enhancer, gene), the model-learned intensities for each epigenomic mark were correlated with mRNA levels of nearby genes. P.C.C.: Pearson correlation coefficient.

Figure S11. 5-hmC decrease in gene bodies and increases in promoters during differentiation. (A) Distribution of 5-hmC signals in different types of genomic features (columns). Fold enrichment: the ratio between the percentage of 5-hmC containing segments in this type (column) and the average percentage of 5-hmC containing segments in the genome (all columns). (B) Average intensities of 5-hmC in 32525 genes. 2,000 bp upstream and 8,000 downstream regions to transcription start site were plotted.

Figure S12. 5-mC does not confound the association of 5-hmC and gene expression. The average signals (normalized by total reads in each dataset) of 5-mC (data from MeDIP-seq) and ^{u}CpG (unmethylated CpG, data from MRE-seq) are shown for promoter groups (A, B) and enhancer groups (C, D). The plots were drawn for $\pm$3000bp regions centering at the centers of all the genomic segments of each group.

Figure S13. Time-course ChIP-seq, MeDIP-seq, MRE-seq and RNA-seq data on a novel Sox 17 enhancer. (A) A GATE predicted mesendoderm enhancer. This enhancer is about 50,000bp upstream of endoderm maker gene *Sox17*. It contains a strong FOXA2 binding site (Figure 5D). Insert: *Sox17* expression is strongly induced during differentiation. (B-C) Increases of H3K4me2/3, 5-hmC, 5-mC and decease of H3K27me3 on this enhancer during differentiation.

Figure S14. FOXA2 and SOX17 ChIP-seq/chip peaks co-localize with predicted enhancers. (A) The GATE predicted enhancer ~50,000 nt upstream of the *Sox17* gene co-localized with a peak (p-value<10^{-6}) in a FOXA2 ChIP-seq experiment in mouse liver (MacIsaac et al. 2010), one of the endoderm-derived organs. The predicted FOXA2 TFBS appeared at the center of this peak. (B) The GATE predicted enhancer ~ 7,000 nt upstream of the *Foxa2* gene co-localized with a strong peak in a SOX17 ChIP-chip experiment in *Sox17*-induced mouse ES cells (Niakan et al. 2010). The predicted SOX17 TFBS located precisely at the center of this peak.

Supplementary tables

Table S1. Data Summary

<table>
<tr><th rowspan="2">Feature</th><th rowspan="2">Mark</th><th rowspan="2">Technology</th><th colspan="3">Differentiation time</th></tr>
<tr><th>Day 0</th><th>Day 4</th><th>Day 6</th></tr>
<tr><td>Repression</td><td>H3K27me3</td><td>ChIP-seq</td><td colspan="3" rowspan="6">(Xiao et al. 2012)</td></tr>
<tr><td rowspan="2">Enhancer</td><td>H3K4me1</td><td>ChIP-seq</td></tr>
<tr><td>H3K27ac</td><td>ChIP-seq</td></tr>
<tr><td rowspan="2">Enhancer/Promoter</td><td>H3K4me2</td><td>ChIP-seq</td></tr>
<tr><td>H2A.Z</td><td>ChIP-seq</td></tr>
<tr><td>Promoter</td><td>H3K4me3</td><td>ChIP-seq</td></tr>
<tr><td>Gene body</td><td>H3K36me3</td><td>ChIP-seq</td></tr>
<tr><td>DNA methylation</td><td>5-hmC</td><td>5-hmC-seq</td><td>√</td><td>√</td><td>√</td></tr>
<tr><td colspan="2">DNA methylation</td><td>MeDIP-seq</td><td rowspan="3">(Xiao et al. 2012)</td><td>√</td><td>√</td></tr>
<tr><td colspan="2">DNA methylation</td><td>MRE-seq</td><td>√</td><td>√</td></tr>
<tr><td colspan="2">Transcriptome</td><td>RNA-seq</td><td colspan="2">(Xiao et al. 2012)</td></tr>
<tr><td colspan="2">Small non-coding RNA</td><td>ncRNA-seq</td><td>√</td><td>√</td><td>√</td></tr>
</table>

√: This study.

Table S2. **Parameters used in simulation.**

Clusters	**Number of segments**	**Markers**	**Transition probability (*b*)**	$\lambda_{0,m}^{k}$	$\lambda_{1,m}^{k}$
Cluster1	2000	M1	From/To 0 1 0 (0.2 0.8) 1 (0 1)	10	20
		M2		10	20
		M3		20	10
		M4		20	10
Cluster2	2000	M1	From/To 0 1 0 (0.5 0.5) 1 (0 1)	20	5
		M2		20	5
		M3		5	20
		M4		5	20
Cluster3	2000	M1	From/To 0 1 0 (0.8 0.2) 1 (0 1)	20	10
		M2		20	10
		M3		10	10
		M4		10	10
Cluster4	2000	M1	From/To 0 1 0 (0.2 0.8) 1 (0 1)	15	50
		M2		25	50
		M3		40	15
		M4		45	10

Table S3. Comparison of model-learned parameters with real parameters in simulation.

Clusters	Markers	Real parameters			Estimated parameters		
		$\lambda_{0,m}^{k}$	$\lambda_{1,m}^{k}$	$b_{0\to0}$	$\lambda_{0,m}^{k}$	$\lambda_{1,m}^{k}$	$b_{0\to0}$
Cluster1	M1	10	20	0.2	11.08634	19.86851	0.1851
	M2	10	20		9.086341	20.01036	
	M3	20	10		19.61862	9.68374	
	M4	20	10		19.50221	9.876021	
Cluster2	M1	20	5	0.5	21.51281	4.945095	0.4995
	M2	20	5		19.48665	5.073476	
	M3	5	20		4.422566	20.07509	
	M4	5	20		4.579186	19.89342	
Cluster3	M1	20	10	0.8	19.4978	10.01112	0.7826
	M2	20	10		20.55367	10.08473	
	M3	10	10		10.03567	10.07011	
	M4	10	10		10.86595	9.911532	
Cluster4	M1	15	50	0.2	15.23131	48.16273	0.2153
	M2	25	50		24.62459	50.68123	
	M3	40	15		41.28035	14.38912	
	M4	45	10		44.31761	9.68048	

Table S4. Comparison of model-learned hidden states with real hidden states in simulation. 0 and 1 correspond to the two hidden activity states. "0->0->0" denotes State 0, 0, 0 in the three time points.

Clusters	Hidden states	Simulated number of segments	Model-inferred number of segments
Cluster1	(0->0->0)	81	78
	(0->0->1)	342	361
	(0->1->1)	1577	1561
Cluster2	(0->0->0)	502	504
	(0->0->1)	511	509
	(0->1->1)	987	987
Cluster3	(0->0->0)	1333	1331
	(0->0->1)	279	282
	(0->1->1)	388	387
Cluster4	(0->0->0)	178	178
	(0->0->1)	411	411
	(0->1->1)	1411	1411

Table S5. Primers for qPCR experiments.

Genes	Primers
Actin_F	AGGCTCTTTTCCAGCCTTCCT
Actin_R	GTCTTTACGGATGTCAACGTCACA
Nanog_F	AGCCTCCAGCAGATGCAAGA
Nanog_R	TGCCACCGCTTGCACTT
Sox2_F	AACTTTTGTCCGAGACCGAGAA
Sox2_R	CCGCGGCCGGTATTTATAAT
Oct4_F	CTGGGCGTTCTCTTTGGAAA
Oct4_R	TATCTCCTGAAGGTTCTCATTGTTG
Gsc_F	AGAACCGCCGAGCCAAGT
Gsc_R	CTCCGGCGAGGCTTTTG
Sox17_F	CCAACACTCCTCCCAAAGTA
Sox17_R	GGTCCAGAATGCATTTTCTC
Chordin_F	TATGCCTTGGACGAGACG
Chordin_R	GGCACTGAGGTTTGATGTTC
Lim1_F	CATATCCGTGAGCAACTGG
Lim1_R	CGCTTAGCTGTTTCATCCTT
Foxa2_F	TTCGAGAACGGCTGCTACCT
Foxa2_R	GGTCTTCTTGCCTCCGCTACT
Cdx2_F	TCACCATCAGGAGGAAAAGT
Cdx2_R	ATTTTCCTCTCCTTGGCTCT
Sox9_F	CGGCTCCAGCAAGAACAAG
Sox9_R	TGCGCCCACACCATGA
Pth1r_F	AGCTCTGGGCACAAGAAGT
Pth1r_R	GTACTGCTGCCTGGTGTCA
Sox7_F	AGGATGAGAGGAAACGTCTG
Sox7_R	TAGGGTCTCTTCTGGGACAG
Hnf4_F	GTGTTTAAGGACGTGCTGCT
Hnf4_R	TCATCAATCTGCAGCTCTTG

Reference

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.

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.

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.