**‘Traffic light rules’: chromatin states direct miRNA-mediated network motifs
running by integrating epigenome and regulatome**

Hongying Zhao^{a,†}, Guanxiong Zhang^{a,†}, Lin Pang^{a,†}, Yujia Lan^{a,†}, Li Wang^a, Fulong Yu^a, Jing Hu^a,
Feng Li^a, Tingting Zhao^c, Yun Xiao^{a,b,*}, Xia Li^{a,*}

*To whom correspondence should be addressed.

†These authors contributed equally to this work.

^a College of Bioinformatics Science and Technology, Harbin Medical University,
Harbin 150081, China.

^b Key Laboratory of Cardiovascular Medicine Research, Harbin Medical University,
Ministry of Education.

^c Department of Neurology, The First Affiliated Hospital of Harbin Medical
University, Harbin, Heilongjiang 150001, China.

***Corresponding Authors:**

Yun Xiao, PhD, College of Bioinformatics Science and Technology, Harbin Medical
University, Harbin 150081, China. Email: xiaoyun@ems.hrbmu.edu.cn.

Xia Li, PhD, College of Bioinformatics Science and Technology, Harbin Medical
University, Harbin 150081, China. Email: lixia@hrbmu.edu.cn. Phone:
86-451-86615922; Fax: 86-451-86615922.

Abstract

Background

Epigenetic marks can cooperatively regulate chromatin accessibility and in turn facilitate or impede the binding of regulatory factors to various elements, suggesting their important roles in regulatory circuits. However, it remains elusive as to how epigenetic marks cooperate in the operations of regulatory network.

Methods

Here, we systematically characterized chromatin states of 26 epigenetic marks on different elements of protein-coding genes and miRNAs. We comprehensively analyzed, by using an integrative regulatory network, how cooperation among epigenetic, transcriptional, and post-transcriptional regulations came about.

Results

We observed extensive cooperation of epigenetic marks on local functional elements and complex epigenetic patterns corresponding to different biological functions. By identifying the significantly epigenetic state-modified motifs, we found that multiple combinations of epigenetic states were associated with a specific type of motif.

Interestingly, miRNA-mediated motifs were linked to stable epigenetic states of downstream targets. Changes in epigenetic states of downstream targets in miRNA-mediated motifs can buffer the effects of upstream regulator on target genes, suggesting that miRNA-mediated motifs require the cooperation of epigenetic marks.

Conclusions

Overall, epigenetic marks are involved in the running of regulatory motifs in the way

traffic lights control traffic flows and hence should be part of the architecture of complex regulatory circuits.

General Significance

We demonstrated a detailed analysis of the cooperation of multiple epigenetic marks and how epigenetic regulation was organized into a human regulatory network. The findings form a basis for further understanding of the complicated roles of epigenetic marks on regulatory circuits.

Highlights

Extensive cooperation of epigenetic marks on local functional elements.

Close association between epigenetic marks and regulatory motifs.

MiRNA-mediated motifs were linked with stable epigenetic states of targets.

Epigenetic states of targets in miRNA-mediated motif can buffer the regulator's effect on target gene expression.

Keywords: miRNA, chromatin states, miRNA-mediated network motifs, epigenetic regulation, transcriptional regulation, post-transcriptional regulation;

1. Introduction

Transcription factors (TFs), critical regulators of gene expression, can combine with other TFs to specify the on-and-off states of genes in a combinatorial fashion [1, 2].

Protein-DNA interactions between TFs and the cis-regulatory target genes (TF-target) constitute a transcriptional regulatory network [3]. MicroRNAs (miRNAs) that allow fine regulation of gene expression at the post-transcriptional level [4] are important partners of TFs. RNA-RNA interactions between miRNAs and the 3'UTR of their

target mRNAs (miRNA-target) constitute a post-transcription regulatory network [3]. TFs and miRNAs contribute to the precise temporal control of gene expression in an interdependent and synergistic way [5]. Thus, the interplay of the two types of regulatory networks forms a complex wiring diagram for a cell [1]. In the complex regulatory networks, the node sets represent molecular components (for example, TFs, genes and miRNAs) and the edge sets describe the regulatory relationships between the nodes [6]. In embryonic stem cells (ESCs), TF and miRNA-involved core regulatory circuitries controlling self-renewal and pluripotency have been described [7]. Several lines of evidence indicates that the complex regulatory network consists of a small set of recurring regulation patterns called network motifs [8]. Network motifs are thought to be recurring interaction circuits which occur in the transcription network more often than would be expected in random networks [8]. Different types of network motifs have been experimentally found to perform specific functions [9] responding to diverse signaling events [10]. Moreover, network motifs have been found to be conserved in various organisms from bacteria to human, indicating that network motifs acting as basic building blocks may exert important effects on the regulatory network [11]. For example, the well-known feed-forward loops (FFLs), in which a master TF regulates a miRNA and both of them target another gene, are frequently overrepresented in mammalian regulatory networks [11]. Through transcriptional and post-transcriptional regulations, FFLs provide relatively stable and consistent activity, and participate in the information-processing function [2, 12].

Several lines of evidence indicate that epigenetic regulation represents another

fundamental mechanism for controlling gene expression [13]. Epigenetics, characterized by stable transmission of heritable information, refers to altered states of gene expression without changes in the DNA sequence [14-16]. Covalent modification of histone tails [17] is a major type of epigenetic mechanism mediating gene expression [18]. For example, the epigenetic mark H3K4me3 (histone H3 trimethylated at lysine 4) is frequently enriched at transcription start sites (TSS) of expressed genes [19]. The active histone mark H3K36me3 (trimethylation of histone H3 on Lys36) is related to transcription elongation and generally covers the whole transcription region of actively expressed genes [20]. The repressive mark H3K27me3 (trimethylation of lysine 27 of histone H3) catalyzed by the polycomb repressor complex 2 can elicit transcriptional repression and therefore is always depleted for active genes [21]. Recent studies indicate that histone modification marks composed of H3K4, H3K9, and H3K27 methylation can influence TF binding [22]. Lupien *et al.* confirmed that removal of methyl groups from H3K4 by demethylase *KDM1* [23] negatively affects *FOXA1* binding but not vice versa, thus reflecting the importance of the H3K4me1,2 marks for *FOXA1* binding [24]. Histone modification cooperates with chromatin remodelers to destabilize local structures of nucleosomes, the basic components of chromatin, for TF binding and subsequent transcriptional activation [25]. H2A.Z, a type of histone variant, can also alter nucleosome properties [26], contributing to accessible chromatin architecture in favor of TF binding. DNase I hypersensitivity can reflect localized open chromatin conformation where DNA is accessible for binding events of transcription factors [27, 28]. Interestingly, in ESCs

the promoters of both *Let-7g* and pluripotency-associated factor *Lin28* showed properly active chromatin without the binding of PcG, a complex catalysing repressive mark H3K27me3, thus allowing the occupation of *Oct4/Sox2/Nanog/Tcf3* and then forming an incoherent feed-forward circuit that contributes to maintaining the pluripotent state [29]. Once differentiation begins, the silence of *Lin28* can relieve *Let-7g* repression [30].

These findings suggest that chromatin marks may be pivotal to dictate where and when TFs bind to their target sites [31]. Furthermore, analysis of histone modifications on gene promoters reveals that genes involved in common functions tend to undergo similar chromatin modifications [32]. Chromatin modification states are like traffic lights when they direct the binding of TFs in a specific condition for maintaining transcriptional circuits that are necessary for the condition [33]. Even though numerous earlier studies focused on the network motifs of biological network at the transcriptional level and the post-transcriptional level [2], it is particularly interesting to ask how epigenetic modification states direct the running of different types of network motifs in regulatory circuits—just like the ‘traffic light rules’ in which colored signaling lights are used to control competing flows of traffic.

In our study, we combined a large number of genome-wide ChIP-seq data (including histone modifications, RNA polymerase II, and histone variant H2A.Z) and DNase I hypersensitivity data in the H1 cell line, to comprehensively characterize the epigenetic modification states of different elements of protein-coding genes and miRNAs. To understand how chromatin states cooperate with transcription regulatory

circuits, we constructed an integrated regulatory network by merging TF-gene and TF-miRNA binding as well as miRNA-gene targeting. By assigning different epigenetic modification states onto the network, a meta-network was built to reflect the combination of epigenetic, transcriptional, and post-transcriptional regulation. We then paid close attention to overrepresented 3-node network motifs together with corresponding chromatin states by comparing them with random networks and comprehensively analyzed how chromatin states indicate network motifs, therefore regulating important biological functions (Figure 1).

2. Materials and Methods

2.1 Epigenome and transcriptome data of the H1 cell line

We obtained data from chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) (composed of H3K4me1, H3K4me2, H3K4me3, H3K36me3, H3K79me1, H3K79me2, H3K9me3, H3K27me3, H4K20me1, H3K23me2, H3K9ac, H3K27ac, H4K91ac, H2AK5ac, H2BK120ac, H2BK15ac, H2BK12ac, H2BK20ac, H2BK5ac, H3K18ac, H3K4ac, H3K56ac, H4K5ac, POL2, and H2A.Z) and DNase I hypersensitivity data in the H1 cell line from NCBI (SRA) database (ID: SRP000941) generated by the Roadmap Epigenomics Project (<http://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics/>). Fifteen marks have multiple replicates. Thirteen of these marks, including H4K91ac, H2AK5ac, H3K4me3, H2BK15ac, POL2, H3K27me3, H2BK12ac, H2BK20ac, H2BK5ac, H3K9me3, H3K9ac, H4K5ac, H3K56ac, contain two biological replicates, and two of them, H3K4me1 and H2BK120ac contain three biological replicates (Supplementary

Fig. S1). In order to reduce the impact of background noise present in individual replicates and therefore to improve sensitivity, biological duplicates were merged before mapping [34, 35].

For ChIP-seq data, original reads were pre-processed using the FASTX-toolkit. Reads were trimmed for the illumina adapter sequence (GATCGGAAGAGCTCGTATGCCGTCTTCT), followed by the trimming of bases with poor quality scores (PHRED score ≤ 2). The trimmed reads with lengths longer than 28 nucleotides were retained. After pre-processing, reads were aligned to the human reference genome (hg18) using the Bowtie algorithm [36], allowing no more than two mismatches. Then, uniquely aligned reads were extended to average length of the DNA fragments (approximately 200 bp) [37-39] in the 3' direction and were used to determine signal intensities of chromatin marks and the hypersensitivity signal (Supplementary Fig. S2).

2.2 Calculating signal intensities of chromatin marks on gene elements

Protein-coding genes were divided into nine regions, namely, promoter (2.5 kb upstream of TSS), 5'UTR, first exon, first intron, internal exon, internal intron, last intron, last exon, and 3'UTR. In parallel, miRNA genes were divided into five regions, namely, promoter (2.5 kb upstream of predicted TSS), 5'region (0.5 kb downstream of predicted TSS), between (from the end of 5'region to the start of pre-miRNA), pre-miRNA, and 3'flank (0.5 bp downstream of the end of pre-miRNA)

(Supplementary Fig. S2). Previous studies have used the 1 bp overlap strategy to calculate the number of reads overlapping a given genomic region [20, 40, 41]. To

calculate signal intensities of a specific chromatin mark for each genomic region, the numbers of extended reads with at least 1 bp [20, 40, 41] overlapping with the genomic region were counted using the corresponding ChIP-seq data and the input data, respectively. When a single read overlaps multiple genomic regions, it is counted for each of these regions. Subsequently, we normalized the read counts by computing the number of reads per kilobase of bin per million reads sequenced (RPKM value), which is widely used to normalize the counts based on genomic length and sequencing depth [38, 40, 42, 43]. The signal intensity i was then calculated by [44, 45]

$$i = \log_2 \left(\frac{R_C + 0.001}{R_I + 0.001} \right) \quad (1)$$

where R_C and R_I represent the RPKM values in ChIP-seq and input data, respectively (Supplementary Fig. S2). Additionally, to investigate the effects of different overlap parameters on the quantification of marks, we calculated the numbers of reads overlapping a given genomic region by a gradient cutoff (ranging from 10% to 100% of the fragment size, at intervals of 10%). The RPKM-normalized values were then used to compute pairwise Pearson correlation coefficients (PCCs) between the 1 bp and the other cutoffs. As a result, the PCC of histone modification H3K4me1 between the 1 bp and 40% cutoffs is up to 0.8. Even compared with the 100% cutoff, the PCC could reach 0.6 (Supplementary Fig. S3A). The average PCCs for all marks range from 0.71 to 0.82, with an average of 0.75, and the average PCCs between the 1 bp and 40% cutoffs is 0.71 (Supplementary Fig. S3B and S4). Our results suggest a comparable effect of different overlap parameters on the quantification of marks. For

genes with multiple internal exons, we calculated the RPKM using the total number of reads overlapping with one of internal exons and the total lengths of all internal exons. Similarly, identical criteria was applied to identify RPKM values for genes with multiple internal introns.

2.3 Calculation of expression levels for protein-coding genes and miRNAs

RNA-seq and smRNA-seq data depicting the expression of protein-coding genes and miRNAs in the H1 cell line were obtained from the NCBI SRA database (ID: SRX007165). For RNA-seq data, transcript assembly and subsequent quantification of expression were performed using cufflinks [46]. RPKM values of protein-coding genes were then used to represent their expression levels. For smRNA-seq data, reads were trimmed to 25 nucleotides and were used for detection of known human miRNAs, which were processed using miRAnalyzer [47] with the default parameters. The number of mapped reads per million (RPM) was used to quantify miRNA expression levels.

2.4 Construction of integrative regulatory network

We collected miRNA-target interactions from multiple miRNA target databases including TargetScan, DIANA-microT, RNAhybrid, RNA22, miRBase Targets, miRanda, and PicTar. To ensure the reliability of miRNA-target interactions, an interaction was considered only if it was present in at least two of the aforementioned databases. The predicted TSSs of miRNA genes were gathered from the miRstart database [48]. The TSSs of protein-coding genes were obtained from UCSC (<http://genome.ucsc.edu>). For both protein-coding genes and miRNAs, their

promoters were defined as 2.5 kb upstream and 0.5 kb downstream of their corresponding TSSs. To predict TFs that act on these promoters, conserved transcription factor binding sites (TFBS) were derived from the ‘TFBS Conserved’ track (hg18 version) of the UCSC genome browser. If a TFBS of a given TF locates in the promoter of a coding gene (or a miRNA), the TF is regarded to regulate the coding gene (or the miRNA). In addition, TF-coding gene regulatory relationships extracted from TRANSFAC (Professional 11.4) were combined with those derived from conserved TFBS-based prediction. By merging TF-gene, TF-miRNA, and miRNA-target, an integrative regulatory network involving TFs, miRNAs, and protein-coding genes, referring to transcriptional and post-transcriptional controls of gene expression, was constructed.

2.5 Clustering analysis of epigenetic profiles

We performed two-dimensional hierarchical clustering on epigenetic profiles of different elements for protein-coding genes and miRNAs. Hierarchical clustering was performed with euclidean distance and complete linkage using ‘hclust’ from R. The clustering analysis was used to group genes and miRNAs according to similar epigenetic patterns on genomic elements. Simultaneously, clustering analysis could identify cooperative patterns of chromatin marks on specific types of genomic elements.

2.6 Selection of cluster number

In order to determine the optimal number of categories, we chose widely used validation measures, including gap statistic (to be maximized) [33, 49], Dunn Indexes

(to be maximized) [50, 51], average silhouette value (to be maximized) [52, 53], Davies-Bouldin Index (to be minimized) [51, 54] and an integrative method COMMUNAL (to be maximized) [55] to evaluate the clustering. Generally, hierarchical clustering is performed for a range of different numbers of clusters ($K=2:30$). The Davies-Bouldin Index and gap statistic are calculated using ‘clusterSim’ and ‘cluster’ package of R, respectively. Dunn Indexes and average silhouette value are calculated using ‘clValid’ package of R. COMMUNAL used Euclidean distance with a ward agglomeration method for epigenetic profiles of genes and miRNAs using the ‘COMMUNAL’ package of R. The parameter setting of COMMUNAL corresponded to clustering algorithms (k-means and hierarchical clustering) and validity metrics (including average silhouette width, gap statistic, Calinski & Harabasz index, connectivity, Dunn index, Dunn index 2).

2.7 Constructing the cooperative network of epigenetic marks at different elements of protein-coding genes and miRNAs

Based on the epigenetic profile of protein-coding genes and miRNAs, pairwise Pearson correlation coefficient (PCC) between epigenetic marks at different genomic elements using ‘cor’ function in R. The significance of the PCC was assessed using the R function ‘cor.test’. Significantly cooperative interactions were determined using $PCC \geq 0.6$ and $P\text{-value} < 0.05$.

2.8 Identification of significant epigenetic state-modified regulatory motifs

We focused on the fully connected regulatory motifs, each composed of three nodes (referring to four compositions: TF-miRNA-gene, TF-TF-gene, TF-miRNA-TF, and

TF-TF-TF) accompanied by their corresponding chromatin states. To evaluate the statistical significance of a given chromatin state-modified regulatory motif, we computed the frequency of this motif in the integrative regulatory networks. Subsequently, 1000 random networks were generated through randomly swapping edges while keeping the in-/out-degree and the composition ratio of connectors for each node. A *P*-value was then computed as the fraction of random networks that have higher (or equal) frequencies than that observed in the real regulatory network. Ultimately, the overrepresented motifs were determined when the following two criteria were met: i) the frequency should be >5 , and ii) the *P*-value should be <0.001 .

2.9 Functional enrichment analysis

Gene Ontology (GO) analysis of genes was performed using DAVID (<http://david.abcc.ncifcrf.gov>). Significantly enriched GO terms were identified using a *P*-value cutoff of 0.001.

3. Results

3.1 Overview of data, processing, and correlation

A total of 26 marks in the embryonic stem cell line H1, containing 10 distinct histone methylations, 13 distinct histone acetylations, RNA polymerase II (POL2), H2A.Z, and DNase I hypersensitivity (Supplementary Table S1), were analyzed, and their corresponding ChIP-seq data referring to 45 samples were extracted from the Roadmap Epigenomics Project [56]. To comprehensively characterize the chromatin pattern on different elements of genes, each protein-coding gene was divided into nine elements, namely, promoter, 5'UTR, first exon, first intron, internal exon, internal

intron, last intron, last exon, and 3'UTR, and each miRNA gene was divided into five elements, namely, promoter, 5' region, between, pre-miRNA, and 3' flank (see Materials and Methods). For all ChIP-seq data, pre-processed reads were mapped to the human reference genome hg18 (Supplementary Table S2), and a total of 663.3 million reads of 28 nt uniquely mapped to the human genome and the log₂ ratio of RPKM between ChIP-seq and control was estimated for each gene element. In addition, RNA-seq and smRNA-seq from the Roadmap Epigenomics Project were used to quantify the expression of protein-coding genes and miRNAs, respectively, in the H1 cell line.

To examine the stability of individual signals between biological replicates, we downsampled the sequence reads to the lowest number of the uniquely mapped reads with duplicates for each of 15 ChIP factors [57]. The pairwise Pearson correlation coefficients (PCC) of signal intensities on each coding gene element were calculated. We found that 66.7% of marks (10 out of 15) showed an average PCC above 0.6, with 4 marks above 0.8 (Supplementary Fig. S5). Previous studies have demonstrate the feasibility of the strategy for merging multiple replicates in the quantification of ChIP factor [58, 59]. Thus, these biological replicates were merged to determine signal intensities of protein-coding genes and miRNAs [58].

3.2 Identification and analysis of chromatin states for protein-coding genes and miRNAs

3.2.1 *Cooperation of chromatin marks at distinct genomic elements*

Chromatin marks can synergistically control gene transcription [60]. Based on the

aforementioned procedure, we constructed epigenetic profiles on different elements for protein-coding genes and miRNAs (Supplementary Table S3 and S4). Each row in the epigenetic profile represents a protein coding gene or a miRNA. Each column represents a continuous quantification of a mark on a genomic element. Through hierarchical clustering of the epigenetic profile of protein-coding genes, epigenetic modifications on the same gene elements are clustered closely together, especially for the same types of modifications, such as acetylation (Figure 2A), implying that chromatin marks tend to be organized in the form of elements. The combination of chromatin marks has been used to identify various genomic elements, such as enhancers [61]. This common pattern of chromatin marks on a specific type of element also supports cooperative mechanisms of multiple modifications for regulation of transcription. Interestingly, histone modifications on the same types of elements were not fully clustered. For example, modifications on promoters were obviously clustered on three different branches: one only contains eleven histone acetylation marks; another contains POL2, H3K4me2, H3K4me3 and DNase I sensitivity without any acetylation marks; the third contains H3K4me1, H3K23me2, H2A.Z, H3K27ac, and H3K9ac. The different compositions of epigenetic modifications on the same type of genomic elements may indicate different active forms in a specific context [61]. In parallel, we also performed a similar analysis through clustering of epigenetic marks on miRNA elements. Like coding genes, similar epigenetic patterns were found in miRNAs, further supporting the cooperation of epigenetic marks (Figure 2B). Furthermore, based on pairwise correlations among

modification marks on different elements, synergy networks on modification marks were built for protein-coding and miRNA genes (Figure 2C and 2D, respectively). There are obviously close connections between epigenetic marks on the same type of elements. In particular, thirteen chromatin modification marks seem to bridge two neighboring elements. For example, POL2 and H3K4me3 on promoter are highly associated with those on the first intron, and, POL2 on promoter was also found to link with H3K4me3 on the first intron, consistent with previous reports that POL2 and H3K4me3 together contributed to transcriptional activity[62].

3.2.2 Dissection of epigenetic states for protein-coding genes and miRNAs

We sought to classify protein-coding genes or miRNAs into different categories based on epigenetic profiles. In order to determine the optimal number of categories, we chose widely used validation measures, including Gap statistic [33, 49], Dunn index (DI) [50, 51], average silhouette value [52, 53] and Davies-Bouldin Index [54], based on hierarchical clustering (see Materials and Methods). These validation measures reflect different aspects of a clustering structure. An integrative method COMMUNAL [55] combines multiple clustering algorithms and validity metrics to determine the optimal number of clusters (Supplementary methods). As a result, for protein-coding genes, the performance curves of gap statistic, Dunn Index and the average silhouette value show the high quality for different numbers of clusters (such as $K=14, 19$ and 25 for gap statistic; $K \geq 11$ or $K=2, 3, 4$ for Dunn Index; $K=2, 3$ and 4 for the average silhouette value). The result of Davies-Bouldin Index shows the local small ‘knee’ at $K=2, 6, 14$ and 25 . Even using the integrative method COMMUNAL,

we did not observe a consistent result. The COMMUNAL output is optimal at $K=2$, and sub-optimal at $K=4, 5, 7$ and 28 (Supplementary Fig. S6). Likewise, for miRNAs, gap statistic, Davies-Bouldin Index, Dunn Index and average silhouette value show optimal clustering at distinct values of K ($K=2, 8, 15$ and 23 for gap statistic; $K=3, 4, 8$ and 20 for Davies-Bouldin Index; $K=6$ for Dunn Index; $K=2, 3, 4$ and 8 for silhouette value). The integrative method COMMUNAL shows optimal at $K=2$, as well as sub-optimal at $K=4$ (Supplementary Fig. S7). Finally, protein-coding genes and miRNAs were divided into twenty-five and eight clusters, respectively (Supplementary Fig. S6 and Fig. S7; the results obtained with $K=14$ for PCGs and $K=4$ for miRNAs in discussion).

Based on the hierarchical clustering described above for protein-coding genes, a total of 18,520 coding genes were divided into 25 groups (Supplementary Table S5). By averaging signal intensities over genes in each group, we observed unique epigenetic modification patterns on sequential gene elements (Figure 3). Using RNA-seq data from the H1 cell line, we then detected expression of genes and found different expression levels corresponding to different epigenetic patterns (Figure 3), which is in line with recent observations [32], further suggesting that numerous epigenetic marks are synergistically responsible for active expression.

Group 1, with the highest expression levels, showed a strong enrichment of transcription initiation mark H3K4me3 and POL2 binding at the 5' part of the whole gene body (including promoter, 5' UTR, and the first exon and intron) and modest enrichment of transcription elongation mark H3K36me3 at the middle of the gene

body, in conjunction with modifications, to a minor extent, of various histone acetylations at the 5' part (Figure 3 and Supplementary Fig. S8). Meanwhile, the repressive marks H3K27me3, H3K9me3, H4K20me1, and H3K23me2 were nearly depleted from the whole gene body but for a few exceptions. Interestingly, groups with consistently low expression also seem to be regulated through different chromatin contexts, three contexts in total (Figure 3 and Supplementary Fig. S8) were found to be linked to low expression: 1) depleted in almost all epigenetic modifications (including active and repressive marks) across the whole gene body, reflecting a highly condensed conformation of chromatin [61, 63], such as group 24; 2) having broad enrichments of H3K9me3 and H3K27me3 across the whole gene body, as in group 25 having the lowest expression level; and 3) showing obvious enrichments of a mass of marks at 5' UTR and first exon but with a striking absence of any marks at promoter, as in group 20. Taken together, these diverse chromatin contexts suggest complex cooperative forms among epigenetic marks, which are required to be responsible for dynamic and variable gene expression [61, 63].

To characterize relationships between chromatin states and functions, we performed the functional enrichments of the protein-coding genes in each group using the DAVID[64] Gene Ontology (GO) annotation tool (P -value < 0.001). Most groups are linked to at least one biological function (Supplementary Fig. S9). Figure 3 shows the top significantly enriched functions for each group. Obviously, genes with different epigenetic patterns are responsible for different biological functions. For example, genes in group 1 are involved in RNA splicing, genes in group 2 are

significantly enriched in mitotic cell cycle while genes in group 16 are associated with cell-cell signaling. Notably, groups 3 and 11, together accounting for about 36.1% of protein-coding genes, were found to be involved in the cell metabolism process and system development, respectively. These two groups show similar epigenetic marks at the 5' part of the whole gene body, but with high differences at the 3' part. In addition, we found that most groups with low expression are associated with various immune defense mechanisms, such as regulation of type I interferon-mediated signaling pathway in group 21, defense response in groups 22 and 23, and immune response in group 25, consistent with the fact that ESCs are considered unrecognizable by the recipient immune defenses, that is, the 'immune-privileged' features of ESCs [65].

Additionally, we found that *POU5F1*, *TCF3*, and *SOX2*, which were previously known to be important for ESC self-renewal and pluripotency, were found to be associated with different chromatin states (located in group 3, 3, and 11, respectively). *POU5F1* and *TCF3* carry strong active marks and less repressive marks. *SOX2* showed the bivalent modification (enrichment of both H3K4me3 and H3K27me3) with additional H3K3me1, H2A.Z, POL2, and numerous acetylations that could loosen histones resulting in facilitating access [66]. These results suggest that multiple epigenetic mechanisms are involved in controlling the expression of pluripotency regulators, which is in line with a recent study [32].

For miRNA genes, eight groups referring to 1028 miRNAs were identified (Figure 4 and Supplementary Table S6). Like in protein-coding genes, different epigenetic patterns of miRNAs appeared to correspond to different expression levels

(shown by different colors from green, yellow to red). Interestingly, it seems that the patterns of epigenetic marks on miRNA genes are different from those in protein-coding genes. For example, group 8, with the lowest expression, presented modest levels of H3K9me3 and slight levels of H3K27me3 on the region harbouring pre-miRNAs, which also showed low occupancy by numerous acetylation marks. This suggests that different types of RNAs may have distinct epigenetic mechanisms for controlling their activities, probably due to their different structures and functions. In addition, we found lower levels of epigenetic modifications on miRNA genes compared with protein-coding genes.

3.3 Assignments of epigenetic state-derived ‘traffic lights’ onto an integrative regulatory network

At present, researchers have begun to realize that gene expression can be affected by the regulations of TFs, miRNAs and epigenetic marks [29]. These regulations from the transcriptional, post-transcriptional, and epigenetic levels form an extremely complex regulatory network that can finely tune expression changes. To construct such an epigenetic state-modified regulatory network, we integrated TF-mediated transcriptional regulations, miRNA-mediated post-transcriptional regulations, and epigenetic mark-mediated regulations.

The predicted targets of miRNAs were combined using seven databases, such as TargetScan (see Materials and Methods). The miRNA-target interactions predicted by at least two algorithms were maintained for further analysis. Briefly, we identified a total of 45,309 predicted miRNA-target interactions, including 507 miRNAs and

7,578 genes. The predicted TSS of miRNA genes were extracted from the miRstart database. Using predicted conserved TFBS track from the UCSC and TransFac database, 2654 TF-miRNA binding interactions referring to 568 miRNAs and 130 TFs, and 44,841 TF-gene interactions referring to 352 TFs and 11,702 genes were determined. By integrating these interactions, a TF-miRNA-gene network containing 92,804 edges and 14,059 nodes composed of 352 TFs and 779 miRNAs as well as 12,928 protein-coding genes was constructed.

We further assessed the accuracy of the TF-miRNA-gene regulatory network using ChIP-seq data, TF knockout data and a well-annotated transcriptional regulatory subnetwork in embryonic stem cell. We obtained ChIP-seq data of 22 TFs and DNase I hypersensitive sites (DHSs) in H1 from ENCODE project. For the TF data, a gene was considered as the target of a TF if at least one binding site of the TF was located within the DHS-containing promoter region of the gene. We observed that 73% of TFs showed significantly overlapping target genes (P -value<0.05, hypergeometric test, Supplementary Fig. S10A). Next, we obtained expression data of small interfering RNA (siRNA) targeting *POU5F1* and *CEBPB* in embryonic stem cells from the GPA database [67]. For each siRNA, differentially expressed genes were identified (fold change >2.5). We found that differentially expressed genes for *POU5F1* and *CEBPB* knockdown were significantly enriched for target genes of *POU5F1* (P -value=4.6e-06) and *CEBPB* (P -value=2.4e-03), respectively (hypergeometric test, Supplementary Fig. S10B). Then, we obtained a core transcriptional regulatory subnetwork, *POU5F1-SOX2-NANOG*, that plays important roles in pluripotency of embryonic

stem cells [68] from Lee *et al.* [69]. As a comparison, we observed nearly identical interactions between our network and the well-annotated core *POU5F1-SOX2-NANOG* subnetwork (Supplementary Fig. S10C). These results suggest the validity of the TF-miRNA-gene regulatory network.

Accumulating evidence indicates that epigenetic modifications function as ‘traffic lights’ to direct the running of regulatory network. To characterize the epigenetic state for each node in the network, a ‘traffic light’ represented by colors based on corresponding epigenetic patterns and transcriptional levels, ranging from green to red in groups 1 to 25 was assigned (Figure 3). Similarly, ‘traffic lights’ were also generated for miRNA groups (Figure 4). The ‘traffic lights’ of TFs, miRNAs, and coding-genes were then mapped onto the integrative regulatory network, forming an epigenetic state-modified regulatory network that represents regulations at transcriptional, post-transcriptional, and epigenetic levels. Importantly, they can help us to understand the ‘traffic light rules’ in biological conditions.

To characterize the association between chromatin state of genes and network structure, we examined topological measures of the regulatory network, including out-degree (O), in-degree (I) and betweenness. We found that TFs with high out-degrees tend to be associated with active chromatin states relative to those with low out-degrees (Supplementary Fig. S11A). However, it was not the case for miRNAs. Furthermore, to assess the diversity of chromatin states of targets, the entropy measure was used [70]. We found that the entropies of TF-regulated PCGs were lower than those of miRNA-regulated PCGs (Supplementary Fig. S11B). It

suggests that the PCGs with TF regulation show relatively stable chromatin states relative to those with miRNA regulation. A ‘hierarchy height’ metric, $h = (O-I)/(O+I)$, was computed to measure the direction of information flow. All nodes were classified into three levels: top-level managers ($0.5 \leq h \leq 1$) regulating many other factors, middle-level managers ($-0.5 < h < 0.5$) which are both regulated and regulating, and bottom-level factors ($-1 \leq h \leq -0.5$) more regulated than regulating [1]. We identified 620 top-level regulators (including 467 miRNAs and 153 TFs), 115 middle-level regulators (including 34 miRNAs and 71 TFs) and 13,844 bottom-level targets (including 278 miRNAs and 13,566 PCGs). When compared with lower-level nodes, upper-level regulators showed higher betweenness, which underlined their central roles in the network architecture, especially for top-level regulators (Supplementary Fig. S11). Moreover, we found that upper-level regulators (especially at middle-level) showed less diversity of chromatin states than bottom-level ones. The compositions of chromatin states are different between upper-level and bottom-level. Top-level regulators are related to strong active chromatin states and bottom-level targets are associated with moderately active states (Supplementary Fig. S12). These results demonstrate that active regulators cooperate with broad epigenetic marks to precisely adjust the transcriptional program under distinct pathophysiological context.

3.4 Traffic light rules: identification of significant epigenetic state-modified regulatory motifs

We searched for fully connected three-node motifs coupled with their corresponding ‘traffic lights’ in the network to study the guidance effects of chromatin marks on

various circuits. Subsequently, 1,000 random networks were generated to determine the significance of epigenetic state-modified motifs (details in Materials and Methods). Finally, seven types of significantly overrepresented motifs were identified with the threshold of P -value less than 0.001 (Figure 5A). It is seen that certain basic types of motifs are more prevalent than would be expected by chance, such as FFLs including both incoherent and coherent types of motifs in which two upstream factors (*i.e.*, TFs and/or miRNAs) are likely to co-regulate their target genes. Besides, complex loops were also identified by combining feed-forward and feedback motifs. To our surprise, we found that almost all types of motifs show multiple compositions of ‘traffic lights,’ except for one type of motif (type 7, termed regulated feedback loop), in which two TFs that regulate each other are regulated by common upstream TF(s) and such motifs contain only one composition of traffic lights. Diverse combinations of traffic lights for each type of motif strongly suggest that epigenetic marks exert important roles in controlling regulatory circuits. However, the type 7 motif shows an unexpected combination of traffic lights, indicating that epigenetic marks do not prefer to participate into controls of this type of motif.

Figure 5B shows all combinations of traffic lights for each type of motif. At first glance, most traffic lights of the two upstream factors are located in the second quadrant, where upstream factors show green or light yellow lights corresponding to relatively open chromatin states and relatively high expression levels. It was noted that downstream target genes also tend to show green or light yellow lights, especially for type 3 and type 4 motifs, both involving a downstream miRNA and two upstream

TFs. Compared with other types of motifs, downstream factors (*i.e.*, miRNAs) in type 3 and type 4 motifs show more consistent traffic lights when upstream TFs exhibit different combinations of lights, even with one of the upstream TFs showing weakly activated epigenetic states. Interestingly, in the type 1 motif, a few compositions of traffic lights of the two upstream factors are located in the third and fourth quadrants (Figure 5B); however, their common downstream target genes show red lights corresponding to repressive chromatin states and low expression levels. Additionally, it is interesting that downstream targets in type 1 (and type 2) motifs showed various types of lights from green to red even when the lights of the two upstream TFs were unchanged. However, the variety was not observed in type 3 (and type 4) motifs, although they share similar structures. This may be explained by a recent report that different signs of regulation may have distinct functions despite the same geometrical structure [2]. The large variety in type 1 and type 2 motifs also supports that epigenetic marks together with upstream regulators co-influence downstream target genes.

Moreover, we found that type 2 motifs have relatively limited combinations of lights compared with type 1 motifs. A similar phenomenon was observed when type 4 and type 3 motifs were compared. Relative to type 1 (or type 3) motifs, type 2 (or 4) motifs add an additional feedback between the two upstream TFs. Previous studies reported that the feedback mechanism can produce more stable states [71]. Our findings further suggest that the additional feedback mechanism is also related to the increased stability of chromatin states of downstream targets.

In addition, two types of miRNA-mediated motifs, namely, type 5 and type 6 motifs, were also identified. Compared with type 1 motif, we found more specialized epigenetic states of downstream target genes in these two types if the lights of the two upstream factors were fixed. There were also a few exceptions in which different epigenetic states of downstream targets were observed when the lights of the two upstream factors were given (Figure 6A and 6B). In such cases, epigenetic marks appeared to play major roles in the control of expression of downstream targets.

Furthermore, we observed that in type 5 motifs, lights of downstream target genes in seven types of FFLs were lighted in the same row but not in the same column. For example, in the condition of upstream miRNAs showing the 4th chromatin pattern, which reflected weak epigenetic marks, when the lights of upstream TFs changed from green to red (Figure 6C), interestingly, the corresponding lights of downstream targets changed from yellow to green (that is, gradually increased active marks and enhancer-associated mark H3K27ac, and decreased repressive marks). These results might suggest that downstream targets could alter chromatin states *per se* to buffer the effects of upstream TFs in order to maintain the expression of downstream target genes, thus supporting that the coordination of chromatin states between regulators and targets might be an important guidance signal to functional motifs. Notably, in type 6 motifs lights of downstream target genes in eight types of FFLs were also lighted in the same column (Figure 5B: type 6). In such conditions, alterations of epigenetic states of miRNAs affected their expression, which in turn regulated downstream targets at a post-transcriptional level. Surprisingly, changes of epigenetic

states of downstream targets also occurred. For example, when the epigenetic modifications of miRNAs were reduced, the lights of downstream targets changed from yellow to green (*i.e.*, increased epigenetic modifications; Figure 6D). These findings indicate the cooperation between epigenetic modifications and post-transcriptional regulations [72].

Moreover, *POU5F1-SOX2-NANOG-TCF3* could promote *Lin28*, which prevented inadvertent activation of *let-7g* and thus provided the stability and robustness of ESC identity [29]. Using epigenome datasets from ENCODE, we found peaks of open chromatin H3K4me3 at the promoter of *Lin28* and transcription elongation mark H3K36me3 along the body without H3K27me3 modification in ESCs, indicating its active expression. In contrast, differentiated cell lines (including GM12878, HMEC, HSMM, HUVEC, NHEK, and NHLF) consistently showed strong enrichment of H3K27me3 without any active marks, indicating transcription repression (Supplementary Fig. S13). This finding reflected that the silence of *Lin28* during differentiation might result from the changes of chromatin marks. Interestingly, the promoter of *Lin28* shows enrichment of the active mark H3K4me3 in the cancer cell lines K562 and HepG2, and of the enhancer marks H3K4me1 and H3K27ac in HepG2, indicating that the ectopic expression of *Lin28* in cancers [73] might be associated with changes in the chromatin pattern of *Lin28*.

4. Discussion

Our study provides a detailed analysis of the cooperation of multiple epigenetic marks and of how epigenetic regulation is organized into a human regulatory network. Our

results highlight the strong cooperative nature of epigenetic marks on different elements of coding genes and miRNAs and reveal a number of guidance rules of epigenetic marks in regulatory networks.

Among extensive cooperation between various chromatin marks, the “bivalent” chromatin pattern associated with both H3K4me3 and H3K27me3 modifications [74], which are thought to poise genes for activation, is widely presented in different epigenetic states of protein-coding genes and miRNAs, including groups 4, 11, 13, 15, and 16 in coding genes and groups 1 and 3 in miRNAs. Groups 11 and 16 in coding genes with relatively low expression show additional chromatin modifications, such as H3K4me1, H4K20me1, H2A.Z, and POL2, which implicates the potential of their losing H3K27me3 and subsequent transcriptional activation during differentiation [75]. These bivalent genes are found to be associated with development and specific cell functions.

Moreover, we found that different types of motifs (Fig. 6E) exhibit different combinations of epigenetic marks, such as TF-mediated FFLs. Notably, compared with general TF-mediated FFLs, miRNA-mediated motifs show relatively stable epigenetic states, probably because the primary role of miRNAs is the fine-tuning of gene expression [76] and miRNA-mediated motifs can enhance the robustness of gene regulation [11]—miRNAs acting like traffic police help to direct information flow, without needing complex epigenetic regulations. Additionally, in type 3 and type 4 motifs, downstream factors (*i.e.*, miRNAs) showed more consistent traffic lights when upstream TFs exhibited different combinations of lights. It might imply that the

transcriptional activity of miRNAs is dependent on the activity of either of the two upstream regulators, rather than on both. Such an OR logic control seems to be highly associated with the open chromatin states at the upstream regions of pre-miRNAs. In the type 1 motif, a few compositions of traffic lights of the two upstream factors and common targets corresponded to repressive chromatin states. It may indicate that, unlike with miRNAs, the transcriptional activity of protein-coding genes is regulated by the activity of both upstream factors. Such an AND logic [77] control of downstream genes also seems to require the cooperation of epigenetic states of downstream genes.

We demonstrated that some general features of the association between chromatin state and network structure. Diversity of chromatin states of TF-regulated PCGs was lower than that of miRNA-regulated PCGs. The compositions of chromatin states are different between upper-level and bottom-level. Next, the upper-level contains information-flow bottlenecks and much connectivity with miRNA. Diversity of chromatin states at upper-level (especially at middle-level) is much lower than that at bottom-level. It perhaps indicated that upper-level regulators (acting as information source) might not require diverse types of chromatin states. However, downstream targets (acting as information receiver) require much more types of chromatin states for finely controlling gene expression. For a type of motif in which two reciprocally regulated TFs are regulated by a common upstream TF, there is only one composition of epigenetic states (active chromatin marks for all three components). Such stable epigenetic states may suggest that this type of motif does not require additional

regulations derived from epigenetic marks. One possible explanation is that tight connections occur in the upper-level of regulatory networks for mitigating information-flow bottlenecks [1]. Stable epigenetic states can maintain the speed of information flow, avoiding the influence of epigenetic marks, just like sustaining a green traffic light for the highest flow movements.

In order to further assess the stability of cooperation between epigenetic marks and operations of regulatory network, based on the evaluation of hierarchical clustering (Supplementary Fig. S6 and Fig. S7; Supplementary methods), protein-coding genes and miRNAs were re-divided into 14 and 4 groups, respectively. Using the clustering results, we re-analyzed the traffic light rules. By averaging signal intensities over genes and miRNAs in each group, we observed unique epigenetic modification patterns on sequential gene elements and different expression levels corresponding to different epigenetic patterns (Supplementary Fig. S14 and Fig. S15). Functional enrichments of the protein-coding genes in each group using the DAVID Gene Ontology (GO) annotation tool (P -value <0.05) showed that genes with different epigenetic patterns are responsible for different biological functions (Supplementary Fig. S14). For example, genes in group 4 are significantly enriched in nuclear division while genes in group 8 are associated with RNA splicing. In addition, we found that most groups with low expression are associated with various immune defense mechanisms, such as immune response in group 12 and defense response in groups 13. Next, by comparing to random networks, seven types of significantly overrepresented motifs were identified (P -value <0.001 ; Supplementary Fig. S16).

Among these significant motifs, a consistent phenomenon is that an additional feedback between the two upstream TFs (type 2 vs. type 1 motifs; type 4 vs. type 3 motifs; Supplementary Fig. S16B) corresponds to relatively limited combinations of lights. Two types of miRNA-mediated motifs, (type 5 and type 6 motifs) were also identified. Comparing type 5 and type 6 motifs with type 1 motif, more specialized epigenetic states of downstream target genes were shown if the lights of the two upstream factors were fixed. In type 5 and type 6 motifs, many lights of downstream target genes were lighted in the same row but not in the same column. These findings exhibited high consistency with ones described in Figure 3, 4 and 5, further supporting cooperation between epigenetic marks and operations of network motifs.

In order to quantify histone modifications in specific regions of the genome and characterize their cooperation, we calculated the Reads per Kilobase per million mapped reads (RPKM) in all regions of interest for ChIP-seq samples [44, 78]. The RPKM normalization corrects biases in sequencing depth across libraries and in gene length within samples. It masks differences in total mRNA content and facilitates transparent comparison of transcript levels both within and between samples [79]. RPKM normalization overcomes the problem of mixed discrete and continuous scales, it enables integrated multi-omics approaches without the requirement of developing specific methods for mixing continuous/discrete scales [80]. However, The RPKM normalization of RNA-seq data has been shown that attempting to correct for differences in gene length in a differential analysis actually has the effect of introducing a bias in the per-gene variances, in particular for lowly expressed genes

[81]. When performing differential analysis of RNA-seq data, RPKM method is ineffective [82]. RPKM normalization method is not able to control false positives in data having a small number of genes with very high read counts [83]. For ChIP-seq data, the widely used normalization approach is based on the sequencing depth (such as RPKM). However, this method may introduce errors when the signal-to-background noise (S/N) ratios vary between samples [84]. It can also be prone to bias caused by unequal variance in different genomic regions [85]. Nonetheless, the RPKM method continues to be a popular choice in many practical applications [82].

In summary, our study highlights the extensive cooperation of chromatin marks and provides a systematic description of the ‘traffic light rules’—how chromatin landscapes indicate the circuitry running in the human regulatory network by integration of transcriptional regulation, post-transcriptional regulation, and epigenetic regulation. The findings form a basis for further understanding the complicated roles of epigenetic marks in regulatory circuits.

Acknowledgements

This work was supported in part by the National High Technology Research and Development Program of China [863 Program, Grant Nos. 2014AA021102], the National Program on Key Basic Research Project [973 Program, Grant Nos. 2014CB910504], the National Natural Science Foundation of China [Grant Nos. 91439117, 61473106, 61170154, 31200997], the National Science Foundation of Heilongjiang Province (Grant Nos. C201207), Wu lien-teh youth science fund project

of Harbin medical university [Grant Nos. WLD-QN1407], and Key Laboratory of Cardiovascular Medicine Research (Harbin Medical University), Ministry of Education. The Undergraduate Innovation Funds of Harbin Medical University [Grant Nos. YJSCX2012-207HLJ, YJSCX2014-21HYD].

Conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] M.B. Gerstein, A. Kundaje, M. Hariharan, S.G. Landt, K.K. Yan, C. Cheng, X.J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A.P. Boyle, P. Cayting, A. Charos, D.Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Fietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O'Geen, Z. Ouyang, E.C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T.E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K.Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P.J. Farnham, R.M. Myers, S.M. Weissman, M. Snyder, Architecture of the human regulatory network derived from ENCODE data, *Nature*, 489 (2012) 91-100.
- [2] C. Cheng, K.K. Yan, W. Hwang, J. Qian, N. Bhardwaj, J. Rozowsky, Z.J. Lu, W. Niu, P. Alves, M. Kato, M. Snyder, M. Gerstein, Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data, *PLoS Comput Biol*, 7 (2011) e1002190.
- [3] N.J. Martinez, A.J. Walhout, The interplay between transcription factors and microRNAs in genome-scale regulatory networks, *BioEssays : news and reviews in molecular, cellular and developmental biology*, 31 (2009) 435-445.
- [4] Y. Xiao, J. Guan, Y. Ping, C. Xu, T. Huang, H. Zhao, H. Fan, Y. Li, Y. Lv, T. Zhao, Y. Dong, H. Ren, X. Li, Prioritizing cancer-related key miRNA-target interactions by integrative genomics, *Nucleic Acids Res*, 40 (2012) 7653-7665.
- [5] J. Chen, G. Wang, C. Lu, X. Guo, W. Hong, J. Kang, J. Wang, Synergetic cooperation of microRNAs with transcription factors in iPS cell generation, *PLoS One*, 7 (2012) e40849.
- [6] B.D. Macarthur, A. Ma'ayan, I.R. Lemischka, Systems biology of stem cell fate and cellular reprogramming, *Nat Rev Mol Cell Biol*, 10 (2009) 672-681.
- [7] V. Kashyap, N.C. Rezende, K.B. Scotland, S.M. Shaffer, J.L. Persson, L.J. Gudas, N.P. Mongan, Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the NANOG, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and stem cell microRNAs, *Stem Cells Dev*, 18 (2009) 1093-1108.
- [8] U. Alon, Network motifs: theory and experimental approaches, *Nat Rev Genet*, 8 (2007) 450-461.
- [9] L. Goentoro, O. Shoval, M.W. Kirschner, U. Alon, The incoherent feedforward loop can provide fold-change detection in gene regulation, *Mol Cell*, 36 (2009) 894-899.
- [10] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, *Science*, 298 (2002) 824-827.

- [11] J. Tsang, J. Zhu, A. van Oudenaarden, MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals, *Mol Cell*, 26 (2007) 753-767.
- [12] M. Kittisopikul, G.M. Suel, Biological role of noise encoded in a genetic network motif, *Proc Natl Acad Sci U S A*, 107 (2010) 13300-13305.
- [13] S.I. Grewal, D. Moazed, Heterochromatin and epigenetic control of gene expression, *Science*, 301 (2003) 798-802.
- [14] G. Lipkin, Plasticity of the cancer cell: implications for epigenetic control of melanoma and other malignancies, *The Journal of investigative dermatology*, 128 (2008) 2152-2155.
- [15] Y.K. Kang, K.K. Lee, Y.M. Han, Reprogramming DNA methylation in the preimplantation stage: peeping with Dolly's eyes, *Current opinion in cell biology*, 15 (2003) 290-295.
- [16] T. Hashimshony, J. Zhang, I. Keshet, M. Bustin, H. Cedar, The role of DNA methylation in setting up chromatin structure during development, *Nat Genet*, 34 (2003) 187-192.
- [17] C.R. Vakoc, S.A. Mandat, B.A. Olenchock, G.A. Blobel, Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin, *Mol Cell*, 19 (2005) 381-391.
- [18] J. Mill, A. Petronis, Molecular studies of major depressive disorder: the epigenetic perspective, *Molecular psychiatry*, 12 (2007) 799-814.
- [19] A. Barski, S. Cuddapah, K. Cui, T.Y. Roh, D.E. Schones, Z. Wang, G. Wei, I. Chepelev, K. Zhao, High-resolution profiling of histone methylations in the human genome, *Cell*, 129 (2007) 823-837.
- [20] T.S. Mikkelsen, M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E.S. Lander, B.E. Bernstein, Genome-wide maps of chromatin state in pluripotent and lineage-committed cells, *Nature*, 448 (2007) 553-560.
- [21] M.D. Young, T.A. Willson, M.J. Wakefield, E. Trounson, D.J. Hilton, M.E. Blewitt, A. Oshlack, I.J. Majewski, ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity, *Nucleic Acids Res*, 39 (2011) 7415-7427.
- [22] M. Shogren-Knaak, H. Ishii, J.M. Sun, M.J. Pazin, J.R. Davie, C.L. Peterson, Histone H4-K16 acetylation controls chromatin structure and protein interactions, *Science*, 311 (2006) 844-847.
- [23] Y. Shi, F. Lan, C. Matson, P. Mulligan, J.R. Whetstone, P.A. Cole, R.A. Casero, Histone demethylation mediated by the nuclear amine oxidase homolog LSD1, *Cell*, 119 (2004) 941-953.
- [24] M. Lupien, J. Eeckhoute, C.A. Meyer, Q. Wang, Y. Zhang, W. Li, J.S. Carroll, X.S. Liu, M. Brown, FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription, *Cell*, 132 (2008) 958-970.
- [25] T. Bartke, M. Vermeulen, B. Xhemalce, S.C. Robson, M. Mann, T. Kouzarides, Nucleosome-interacting proteins regulated by DNA and histone methylation, *Cell*, 143 (2010) 470-484.
- [26] C. Jin, C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao, G. Felsenfeld, H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions, *Nat Genet*, 41 (2009) 941-945.
- [27] S. Neph, J. Vierstra, A.B. Stergachis, A.P. Reynolds, E. Haugen, B. Vernot, R.E. Thurman, S. John, R. Sandstrom, A.K. Johnson, M.T. Maurano, R. Humbert, E. Rynes, H. Wang, S. Vong, K. Lee, D. Bates, M. Diegel, V. Roach, D. Dunn, J. Neri, A. Schafer, R.S. Hansen, T. Kutayavin, E. Giste, M. Weaver, T. Canfield, P. Sabo, M. Zhang, G. Balasundaram, R. Byron, M.J. MacCoss, J.M. Akey, M.A. Bender, M. Groudine, R. Kaul, J.A. Stamatoyannopoulos, An expansive human regulatory lexicon encoded in transcription factor footprints, *Nature*, 489 (2012) 83-90.

- [28] C.G. Spilianakis, R.A. Flavell, Epigenetic regulation of *Ifng* expression, *Nature immunology*, 8 (2007) 681-683.
- [29] A. Marson, S.S. Levine, M.F. Cole, G.M. Frampton, T. Brambrink, S. Johnstone, M.G. Guenther, W.K. Johnston, M. Wernig, J. Newman, J.M. Calabrese, L.M. Dennis, T.L. Volkert, S. Gupta, J. Love, N. Hannett, P.A. Sharp, D.P. Bartel, R. Jaenisch, R.A. Young, Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells, *Cell*, 134 (2008) 521-533.
- [30] J.E. Babiarz, R. Blelloch, Small RNAs - their biogenesis, regulation and function in embryonic stem cells, (2008).
- [31] K.S. Zaret, J.S. Carroll, Pioneer transcription factors: establishing competence for gene expression, *Genes Dev*, 25 (2011) 2227-2241.
- [32] J.A. Wamstad, J.M. Alexander, R.M. Truty, A. Shrikumar, F. Li, K.E. Eilertson, H. Ding, J.N. Wylie, A.R. Pico, J.A. Capra, G. Erwin, S.J. Kattman, G.M. Keller, D. Srivastava, S.S. Levine, K.S. Pollard, A.K. Holloway, L.A. Boyer, B.G. Bruneau, Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage, *Cell*, 151 (2012) 206-220.
- [33] H. Zhao, T. Liu, L. Liu, G. Zhang, L. Pang, F. Yu, H. Fan, Y. Ping, L. Wang, C. Xu, Y. Xiao, X. Li, Chromatin states modify network motifs contributing to cell-specific functions, *Scientific reports*, 5 (2015) 11938.
- [34] H. Lam, E.W. Deutsch, J.S. Eddes, J.K. Eng, S.E. Stein, R. Aebersold, Building consensus spectral libraries for peptide identification in proteomics, *Nature methods*, 5 (2008) 873-875.
- [35] K. Robasky, N.E. Lewis, G.M. Church, The role of replicates for error mitigation in next-generation sequencing, *Nat Rev Genet*, 15 (2014) 56-62.
- [36] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol*, 10 (2009) R25.
- [37] L. Arrigoni, A.S. Richter, E. Betancourt, K. Bruder, S. Diehl, T. Manke, U. Bonisch, Standardizing chromatin research: a simple and universal method for ChIP-seq, *Nucleic Acids Res*, (2015).
- [38] R.D. Hawkins, G.C. Hon, L.K. Lee, Q. Ngo, R. Lister, M. Pelizzola, L.E. Edsall, S. Kuan, Y. Luu, S. Klugman, J. Antosiewicz-Bourget, Z. Ye, C. Espinoza, S. Agarwahl, L. Shen, V. Ruotti, W. Wang, R. Stewart, J.A. Thomson, J.R. Ecker, B. Ren, Distinct epigenomic landscapes of pluripotent and lineage-committed human cells, *Cell Stem Cell*, 6 (2010) 479-491.
- [39] H. O'Geen, L. Echipare, P.J. Farnham, Using ChIP-seq technology to generate high-resolution profiles of histone modifications, *Methods in molecular biology*, 791 (2011) 265-286.
- [40] S.J. Lim, P.J. Boyle, M. Chinen, R.K. Dale, E.P. Lei, Genome-wide localization of exosome components to active promoters and chromatin insulators in *Drosophila*, *Nucleic Acids Res*, 41 (2013) 2963-2980.
- [41] J. Cotney, J. Leng, S. Oh, L.E. DeMare, S.K. Reilly, M.B. Gerstein, J.P. Noonan, Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb, *Genome Res*, 22 (2012) 1069-1080.
- [42] N. Rajagopal, W. Xie, Y. Li, U. Wagner, W. Wang, J. Stamatoyannopoulos, J. Ernst, M. Kellis, B. Ren, RFECs: a random-forest based algorithm for enhancer identification from chromatin state, *PLoS Comput Biol*, 9 (2013) e1002968.
- [43] T.J. Treangen, S.L. Salzberg, Repetitive DNA and next-generation sequencing: computational challenges and solutions, *Nat Rev Genet*, 13 (2012) 36-46.
- [44] G.C. Hon, R.D. Hawkins, O.L. Caballero, C. Lo, R. Lister, M. Pelizzola, A. Valsesia, Z. Ye, S. Kuan, L.E. Edsall, A.A. Camargo, B.J. Stevenson, J.R. Ecker, V. Bafna, R.L. Strausberg, A.J. Simpson, B. Ren, Global

DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer, *Genome Res*, 22 (2012) 246-258.

[45] L. Chen, G. Wu, H. Ji, hmChIP: a database and web server for exploring publicly available human and mouse ChIP-seq and ChIP-chip data, *Bioinformatics*, 27 (2011) 1447-1448.

[46] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, *Nat Biotechnol*, 28 (2010) 511-515.

[47] M. Hackenberg, N. Rodriguez-Ezpeleta, A.M. Aransay, miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments, *Nucleic Acids Res*, 39 (2011) W132-138.

[48] C.H. Chien, Y.M. Sun, W.C. Chang, P.Y. Chiang-Hsieh, T.Y. Lee, W.C. Tsai, J.T. Horng, A.P. Tsou, H.D. Huang, Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data, *Nucleic Acids Res*, 39 (2011) 9345-9356.

[49] D. Grun, A. Lyubimova, L. Kester, K. Wiebrands, O. Basak, N. Sasaki, H. Clevers, A. van Oudenaarden, Single-cell messenger RNA sequencing reveals rare intestinal cell types, *Nature*, 525 (2015) 251-255.

[50] J. Handl, J. Knowles, D.B. Kell, Computational cluster validation in post-genomic data analysis, *Bioinformatics*, 21 (2005) 3201-3212.

[51] R. Guthke, U. Moller, M. Hoffmann, F. Thies, S. Topfer, Dynamic network reconstruction from gene expression data applied to immune response during bacterial infection, *Bioinformatics*, 21 (2005) 1626-1634.

[52] R.G. Verhaak, K.A. Hoadley, E. Purdom, V. Wang, Y. Qi, M.D. Wilkerson, C.R. Miller, L. Ding, T. Golub, J.P. Mesirov, G. Alexe, M. Lawrence, M. O'Kelly, P. Tamayo, B.A. Weir, S. Gabriel, W. Winckler, S. Gupta, L. Jakkula, H.S. Feiler, J.G. Hodgson, C.D. James, J.N. Sarkaria, C. Brennan, A. Kahn, P.T. Spellman, R.K. Wilson, T.P. Speed, J.W. Gray, M. Meyerson, G. Getz, C.M. Perou, D.N. Hayes, N. Cancer Genome Atlas Research, Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1, *Cancer cell*, 17 (2010) 98-110.

[53] D.J. Weisenberger, K.D. Siegmund, M. Campan, J. Young, T.I. Long, M.A. Faasse, G.H. Kang, M. Widschwendter, D. Weener, D. Buchanan, H. Koh, L. Simms, M. Barker, B. Leggett, J. Levine, M. Kim, A.J. French, S.N. Thibodeau, J. Jass, R. Haile, P.W. Laird, CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer, *Nat Genet*, 38 (2006) 787-793.

[54] V.K. Rajasekhar, A. Viale, N.D. Socci, M. Wiedmann, X. Hu, E.C. Holland, Oncogenic Ras and Akt signaling contribute to glioblastoma formation by differential recruitment of existing mRNAs to polysomes, *Mol Cell*, 12 (2003) 889-901.

[55] T.E. Sweeney, A.C. Chen, O. Gevaert, Combined Mapping of Multiple clUsteriNg ALgorithms (COMMUNAL): A Robust Method for Selection of Cluster Number, K, *Scientific reports*, 5 (2015) 16971.

[56] B.E. Bernstein, J.A. Stamatoyannopoulos, J.F. Costello, B. Ren, A. Milosavljevic, A. Meissner, M. Kellis, M.A. Marra, A.L. Beaudet, J.R. Ecker, P.J. Farnham, M. Hirst, E.S. Lander, T.S. Mikkelsen, J.A. Thomson, The NIH Roadmap Epigenomics Mapping Consortium, *Nat Biotechnol*, 28 (2010) 1045-1048.

[57] N. Koike, S.H. Yoo, H.C. Huang, V. Kumar, C. Lee, T.K. Kim, J.S. Takahashi, Transcriptional architecture and chromatin landscape of the core circadian clock in mammals, *Science*, 338 (2012) 349-354.

- [58] M.M. Hoffman, O.J. Buske, J. Wang, Z. Weng, J.A. Bilmes, W.S. Noble, Unsupervised pattern discovery in human chromatin structure through genomic segmentation, *Nature methods*, 9 (2012) 473-476.
- [59] Q. Wang, J. Huang, H. Sun, J. Liu, J. Wang, Q. Wang, Q. Qin, S. Mei, C. Zhao, X. Yang, X.S. Liu, Y. Zhang, CR Cistrome: a ChIP-Seq database for chromatin regulators and histone modification linkages in human and mouse, *Nucleic Acids Res*, 42 (2014) D450-458.
- [60] W.S. Lo, L. Duggan, N.C. Emre, R. Belotserkovskaya, W.S. Lane, R. Shiekhattar, S.L. Berger, Snf1--a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription, *Science*, 293 (2001) 1142-1146.
- [61] J. Ernst, M. Kellis, Discovery and characterization of chromatin states for systematic annotation of the human genome, *Nat Biotechnol*, 28 (2010) 817-825.
- [62] Z. Wang, C. Zang, J.A. Rosenfeld, D.E. Schones, A. Barski, S. Cuddapah, K. Cui, T.Y. Roh, W. Peng, M.Q. Zhang, K. Zhao, Combinatorial patterns of histone acetylations and methylations in the human genome, *Nat Genet*, 40 (2008) 897-903.
- [63] W. Wu, Y. Cheng, C.A. Keller, J. Ernst, S.A. Kumar, T. Mishra, C. Morrissey, C.M. Dorman, K.B. Chen, D. Drautz, B. Giardine, Y. Shibata, L. Song, M. Pimkin, G.E. Crawford, T.S. Furey, M. Kellis, W. Miller, J. Taylor, S.C. Schuster, Y. Zhang, F. Chiaromonte, G.A. Blobel, M.J. Weiss, R.C. Hardison, Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration, *Genome Res*, 21 (2011) 1659-1671.
- [64] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat Protoc*, 4 (2009) 44-57.
- [65] F. Fandrich, B. Dresske, M. Bader, M. Schulze, Embryonic stem cells share immune-privileged features relevant for tolerance induction, *J Mol Med (Berl)*, 80 (2002) 343-350.
- [66] M. Ptashne, A. Gann, Imposing specificity by localization: mechanism and evolvability, *Curr Biol*, 8 (1998) R897.
- [67] Y. Xiao, Y. Gong, Y. Lv, Y. Lan, J. Hu, F. Li, J. Xu, J. Bai, Y. Deng, L. Liu, G. Zhang, F. Yu, X. Li, Gene Perturbation Atlas (GPA): a single-gene perturbation repository for characterizing functional mechanisms of coding and non-coding genes, *Scientific reports*, 5 (2015) 10889.
- [68] Y.H. Loh, J.H. Ng, H.H. Ng, Molecular framework underlying pluripotency, *Cell cycle*, 7 (2008) 885-891.
- [69] T.I. Lee, R.G. Jenner, L.A. Boyer, M.G. Guenther, S.S. Levine, R.M. Kumar, B. Chevalier, S.E. Johnstone, M.F. Cole, K. Isono, H. Koseki, T. Fuchikami, K. Abe, H.L. Murray, J.P. Zucker, B. Yuan, G.W. Bell, E. Herbolsheimer, N.M. Hannett, K. Sun, D.T. Odom, A.P. Otte, T.L. Volkert, D.P. Bartel, D.A. Melton, D.K. Gifford, R. Jaenisch, R.A. Young, Control of developmental regulators by Polycomb in human embryonic stem cells, *Cell*, 125 (2006) 301-313.
- [70] V. Amin, R.A. Harris, V. Onuchic, A.R. Jackson, T. Charnecki, S. Paithankar, S. Lakshmi Subramanian, K. Riehle, C. Coarfa, A. Milosavljevic, Epigenomic footprints across 111 reference epigenomes reveal tissue-specific epigenetic regulation of lincRNAs, *Nature communications*, 6 (2015) 6370.
- [71] J. Li, X. Hua, M. Haubrock, J. Wang, E. Wingender, The architecture of the gene regulatory networks of different tissues, *Bioinformatics*, 28 (2012) i509-i514.
- [72] K.E. Szulwach, X. Li, R.D. Smrt, Y. Li, Y. Luo, L. Lin, N.J. Santistevan, W. Li, X. Zhao, P. Jin, Cross talk between microRNA and epigenetic regulation in adult neurogenesis, *J Cell Biol*, 189 (2010) 127-141.
- [73] S.R. Viswanathan, J.T. Powers, W. Einhorn, Y. Hoshida, T.L. Ng, S. Toffanin, M. O'Sullivan, J. Lu, L.A. Phillips, V.L. Lockhart, S.P. Shah, P.S. Tanwar, C.H. Mermel, R. Beroukhim, M. Azam, J. Teixeira, M.

- Meyerson, T.P. Hughes, J.M. Llovet, J. Radich, C.G. Mullighan, T.R. Golub, P.H. Sorensen, G.Q. Daley, Lin28 promotes transformation and is associated with advanced human malignancies, *Nat Genet*, 41 (2009) 843-848.
- [74] B.E. Bernstein, T.S. Mikkelsen, X. Xie, M. Kamal, D.J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S.L. Schreiber, E.S. Lander, A bivalent chromatin structure marks key developmental genes in embryonic stem cells, *Cell*, 125 (2006) 315-326.
- [75] K. Cui, C. Zang, T.Y. Roh, D.E. Schones, R.W. Childs, W. Peng, K. Zhao, Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation, *Cell Stem Cell*, 4 (2009) 80-93.
- [76] E.M. Small, E.N. Olson, Pervasive roles of microRNAs in cardiovascular biology, *Nature*, 469 (2011) 336-342.
- [77] G. Yao, C. Tan, M. West, J.R. Nevins, L. You, Origin of bistability underlying mammalian cell cycle entry, *Mol Syst Biol*, 7 (2011) 485.
- [78] R. Rebollo, M.M. Karimi, M. Bilenky, L. Gagnier, K. Miceli-Royer, Y. Zhang, P. Goyal, T.M. Keane, S. Jones, M. Hirst, M.C. Lorincz, D.L. Mager, Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms, *PLoS genetics*, 7 (2011) e1002301.
- [79] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq, *Nature methods*, 5 (2008) 621-628.
- [80] M. Hur, A.A. Campbell, M. Almeida-de-Macedo, L. Li, N. Ransom, A. Jose, M. Crispin, B.J. Nikolau, E.S. Wurtele, A global approach to analysis and interpretation of metabolic data for plant natural product discovery, *Natural product reports*, 30 (2013) 565-583.
- [81] A. Oshlack, M.J. Wakefield, Transcript length bias in RNA-seq data confounds systems biology, *Biology direct*, 4 (2009) 14.
- [82] M.A. Dillies, A. Rau, J. Aubert, C. Hennequet-Antier, M. Jeanmougin, N. Servant, C. Keime, G. Marot, D. Castel, J. Estelle, G. Guernec, B. Jagla, L. Jouneau, D. Laloe, C. Le Gall, B. Schaeffer, S. Le Crom, M. Guedj, F. Jaffrezic, C. French StatOmique, A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis, *Brief Bioinform*, 14 (2013) 671-683.
- [83] Y. Lin, K. Golovkina, Z.X. Chen, H.N. Lee, Y.L. Negron, H. Sultana, B. Oliver, S.T. Harbison, Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual *Drosophila melanogaster*, *BMC Genomics*, 17 (2016) 28.
- [84] Z. Shao, Y. Zhang, G.C. Yuan, S.H. Orkin, D.J. Waxman, MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets, *Genome Biol*, 13 (2012) R16.
- [85] C. Taslim, J. Wu, P. Yan, G. Singer, J. Parvin, T. Huang, S. Lin, K. Huang, Comparative study on ChIP-seq data: normalization and binding pattern characterization, *Bioinformatics*, 25 (2009) 2334-2340.

Figure Legends

Figure 1. An overview of the workflow to identify epigenetic state-modified regulatory motifs. The overall procedure includes five major steps: (1) Collecting ChIP-seq data (including 13 distinct histone acetylations, 10 histone methylations, RNA Pol II bindings, DNase I hypersensitivity, and histone variant H2A.Z) in human ESCs H1. (2) Protein-coding genes and miRNA genes were divided into nine and five genomic elements, respectively. For each genomic element, the signal intensities of twenty-six types of marks were calculated separately. Thus, an epigenetic profile was used to characterize epigenetic modification patterns at different genomic elements for protein-coding genes and miRNAs, respectively. The columns of matrixes represent signal intensity ($\log_2(\text{ChIP RPKM}/\text{input RPKM})$) on different elements, and the rows represent protein-coding genes or miRNAs. Each row in the epigenetic profile corresponds to a protein coding gene or a miRNA. Each column represents continuous quantification of a mark on a genomic element. A hierarchical clustering method was used to group genes and miRNAs according to the similarity of epigenetic profile, respectively. Each group of genes represents a type of chromatin states. (3) Constructing a regulatory network by integrating TF-gene, TF-miRNA, and miRNA-gene regulatory relationships. Each red square represents a TF, each green triangle represents a miRNA, and each circle represents a protein-coding gene. (4) Assigning chromatin states of genes to each node in the regulatory network, an integrative epigenetic state-containing regulatory network was constructed. (5) Compared to random networks, we then detected over-represented three-node

epigenetic state-modified network motifs.

Figure 2. Cooperation of chromatin marks at distinct genomic elements.

Hierarchical clustering dendrograms demonstrate relationships between each mark at each genomic element (columns in epigenetic profile) of protein-coding genes (A) and miRNAs (B). Black rectangles refer to three main branches of promoters. The presence of a specific chromatin mark or a genomic element is indicated by a color-coded rectangle. Cooperative networks of epigenetic marks at different elements of protein-coding genes. (C) and miRNA genes (D). Significantly cooperative interactions were determined using the Pearson correlation coefficient ($PCC \geq 0.6$) and the respective *P*-value (*P*-value < 0.05). The node colors indicate the type of genomic elements of coding-genes from (A) and miRNAs from (B).

Figure 3. The signal distribution over each chromatin state of protein-coding genes. For each chromatin state, mean signal distribution of chromatin marks for each genomic element was showed in the heatmap and shaded area. The x-axis represents different genomic elements, and the y-axis represents different epigenetic marks (arranged according to orders in the legend inside the circle). Inside the circle, the box-plot shows the distribution of expression levels of protein-coding genes ($\log_2$ RPKM value) across each chromatin state. Then, the ‘traffic lights’ from green, yellow, to red were assigned to groups 1 to 25 in the clockwise direction based on their chromatin states and transcription levels. Marginal labels depict the top biological functions significantly enriched by protein-coding genes in each group using DAVID (*P*-value < 0.001).

Figure 4. The signal distribution over each chromatin state of miRNAs. (A) The average signal distribution of chromatin marks for each genomic element of miRNAs was shown in the heatmap and shaded area. (B) The box-plot shows the distribution of

expression levels of miRNAs (log2 RPM value) across eight types of chromatin states of miRNAs.

Figure 5. The landscape of epigenetic state-modified motifs. (A) Seven overrepresented structures of epigenetic state-modified motifs are identified with P -value <0.001 . Each red square represents a TF, each green triangle represents a miRNA, and each circle represents a protein-coding gene. (B) The presence of a specific chromatin state composition associated with each type of network motifs. The horizontal axis and vertical axis denote traffic lights of two upstream regulators, and traffic lights inside the regions represent the chromatin states of downstream targets. A circle with multiple divisions indicated that downstream targets showed various types of chromatin states (traffic lights) when the lights of the two upstream TFs were unchanged. Different chromatin states of genes are indicated by color-coded circles. Gray-filled circles represent data not available.

Figure 6. Examples of miRNA-mediated epigenetic state-modified motifs.

(A-C) Different chromatin state compositions associated with a specific type of miRNA-mediated network motif. Each embedded heatmap shows the distributions of chromatin marks corresponding to a specific traffic light at different genomic elements. The rows stand for elements of coding genes or miRNA. The columns stand for mean signal intensities of chromatin marks on each genomic element.

‘H3K4me1_2K79me1_2’ refers to the average signal intensities of H3K4me1, H3K4me2, H3K79me1 and H3K79me2. ‘Acetylation’ refers to the mean of signal intensities of acetylations except for H3K27ac. (E) Three types of traffic rules: (E-1)

The motif that contains TF and targets exhibits diverse combinations of ‘traffic lights’ characterized by epigenetic marks. The motif seems to be under the complex epigenetic regulations; (E-2) The miRNA-mediated FFL shows relatively stable epigenetic states. It seems that miRNAs act like traffic police helping to direct information flow, without needing complex epigenetic regulations; (E-3) A type of motif in which two reciprocally regulated TFs are regulated by a common upstream TF, there is only one composition of epigenetic states, suggesting that this type of motif does not require additional regulations derived from epigenetic marks, just like sustaining a green traffic light for the highest flow movements.

Supplementary information

Supplementary methods; Supplementary Discussion; Supplementary Tables S1-S6; Supplementary Figures S1-S16.

2 Identifying chromatin states of protein-coding genes and miRNAs

3 Constructing integrative regulatory network

4 Constructing an integrative epigenetic state-modified regulatory network

5 Identifying significant epigenetic state-modified network motifs

Figure

C

D

Figure

Figure

B

B

C

D

E-1

E-2

E-3