Mysteries of gene regulation: Promoters are not the sole triggers of gene expression

Chi-Nga Chowa, Kuan-Chieh Tsengb, Ping-Fu Houc, Nai-Yun Wua, Tzong-Yi Leed, Wen-Chi Changd,a

a Institute of Tropical Plant Sciences and Microbiology, College of Biosciences and Biotechnology, National Cheng Kung University, Tainan 70101, Taiwan
b Department of Life Sciences, College of Biosciences and Biotechnology, National Cheng Kung University, Tainan 70101, Taiwan
c Kaohsiung District Agricultural Research and Extension Station, Pingtung County 90846, Taiwan
d School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China

1. Introduction

Cis-regulatory elements located at promoter regions are usually the focal point when studying gene regulation by transcription factors (TFs) [1], but in recent studies, TFs were also observed to bind to 5′ UTRs and introns [2,3], suggesting that the cis-regulatory elements of promoters cannot sufficiently explain the entire mechanism underlying TF regulation. Similarly, studies have revealed that TFs mediate the transcription of non-coding genes such as microRNA and long non-coding RNA (lncRNA) [4–6]. However, this phenomenon in plants was only reported recently; hence, the involvement of TFs and non-coding genes in this regulatory mechanism remains unclear. Accordingly, comprehensive genomic maps were required to explore the regulatory roles of plant TFs and to characterize the cis-regulatory regions of protein-coding genes and non-coding genes.

The rapid accumulation of high-throughput sequencing datasets and the improvement of computational methods have allowed for new insights into the transcriptional regulations of plant genomes. For example, chromatin immunoprecipitation sequencing (ChIP-seq) not only provided genome-wide binding profiles of TFs but was also used as a true positive set to create features in TF binding site prediction tools [7,8]. ChIP-seq coupled with an antibody for detecting histone marks also enabled an increased understanding of the epigenetic regulation during different developmental stages and stress responses [9–11]. Moreover, DNase I hypersensitive sites (DHSs) indicate the genomic regions of chromatin accessible to TF binding for gene activation. Unlike ChIP-seq, the use of DHSs is not limited to examining the binding sites of one individual TF [12]. Notably, emerging chromosome conformation capture-based technologies, such as Hi-C, can define topologically associating domains (TADs), which are the regions of chromatin with high self-interactions. TADs and interactions between promoters and enhancers allow for the prediction of associations between the expression and regulation of genes [13,14]. Although multiple high-throughput sequencing methods have been applied to elucidate gene regulation, the studies applying them have usually narrowed the possible relevant genes/regulators...
down to a group of specific genes or a small number of regulators. Thus, a whole-genome view of gene regulation in plants is absent from the literature.

In this study, we explored the regulatory regions of protein-coding and non-coding genes of TFs and histone marks by using public ChIP-seq datasets. The genome-wide landscapes of TF binding peaks obtained from protein-coding genes and non-coding genes revealed that the genetic regions could vary according to the individual TF. For protein-coding genes, the cis-regulatory regions around both transcription start sites (TSSs) and transcription termination sites (TTTs) generally contained the most TF binding sites. Conversely, the exons of non-coding genes were more vital for their transcriptional regulation than were those of other regions. The histone marks demonstrated that the diverse combinations of histone variants and modifications were used to pack the promoters (or gene bodies) of different gene types. The integration of Hi-C maps and ChIP-seq depictions revealed that TAD boundaries were colocalized with the regions related to gene activation and TF binding. Additionally, the comparisons between non-responsive (NR) and heat-stress-responsive (HS) genes suggested that these two gene sets were substantially different in cis-regulatory regions, histone regulation, and TAD boundary organization. Overall, these results demonstrated the complexity of gene regulation and constituted a worthwhile investigation for integrating the multiple high-throughput sequencing data.

2. Materials and methods

2.1. Extraction of TF and histone deposition preferences from ChIP-seq data

ChIP-seq-based genomic landscapes were retrieved from our database [15]. Samples of 53 TFs and 19 histone marks are listed in Supplementary Tables S1 and S2, respectively. To estimate the distribution of TF binding peaks across the Arabidopsis thaliana genome, the genome sequence annotation file (GFF) was downloaded from the TAIR database (Araport11 version) [16]. A total of 27,445 protein-coding genes and 41,642 non-coding genes were recorded in the GFF file. The subdivided genetic regions of protein-coding and non-coding genes are illustrated in Supplementary Fig. S1. BEDTools was compiled to overlap the genomic features (i.e., gene types and the subdivided genetic regions) with TF binding peaks and histone mark deposition [17]. To prevent overestimating the genomic features within genomic coordinates containing high gene density, the frequency scores of each genomic feature were normalized by numbers of transcripts and genes as per the following formula:

\[ S_r = \frac{P \sum_{g=1}^{C} \sum_{t=1}^{T} \left( \frac{\text{Len}_t}{\text{Len}_t + N_p + N_g} \right) }{ \text{Sr} } \]

where \( S_r \) is the score of one type \( r \) of genomic feature (e.g., the exon), \( P \) is the number of peaks overlapping with type \( r \), \( G \) and \( T \) are the numbers of genes and transcripts overlapping with peak \( p \), respectively, \( \text{Len}_t \) is the occupied region length of peak \( p \), \( \text{Len}_t \) is the length of the overlapping region between type \( r \) of transcripts \( t \) and peak \( p \), \( N_t \) is the number of genes of \( g \) overlapping with peak \( p \), and \( N_g \) is the number of transcripts \( t \) belonging to gene \( g \). The sum score of all genomic features for one peak was 1. The average number of samples for each individual TF/histone mark was calculated for each genomic feature. Finally, Highcharts (https://www.highcharts.com/) was used to visualize the proportion of peaks located at the genomic features.

To construct the distribution of TF binding peaks and histone mark depositions, 5-kb flanking regions of TSSs/TTSSs and ChIP-seq peaks were overlapped by using BEDTools [17]. The upstream and downstream 5-kb flanking regions were divided into nonoverlapping 100-bp windows. The number of overlapping peaks was calculated in each window, from which was subtracted the average of all windows, which yielded the result used to determine the z-score. Specifically, the result was divided by the standard deviation of all windows to yield the z-score. These scores were calculated for each sample. The samples of one individual TF/histone mark were merged by calculating the average.

2.2. GO functional enrichment analysis

To infer the biological processes, molecular function, cellular component, and metabolic pathways of gene sets (i.e., genes regulated by only one genetic region and genes inside/outside the TAD boundaries), the KEGG/GO enrichment analysis function of EXPat 2.0 was used [18]. The cumulative probability \( p \) value of hypergeometric distribution was calculated to evaluate the overrepresented metabolic pathways/GO terms. A \( p \) value of \(<0.05\) was considered statistically significant. Because of the large number of genes inside and outside the TAD boundaries, the cut-off for the false discovery rate (0.1) was used to select enriched GO terms.

2.3. Collection and processing of Hi-C data

The Arabidopsis Hi-C data (20 samples) were collected from the Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA) [19,20]. FASTX-Toolkit (version 0.0.13, http://hannonlab.cshl.edu/~fastx_toolkit/) was used to remove low-quality reads (60% sequences of reads \( \geq \) Q30 and read length \(<30\) bp). Hi-CPro was applied to filter the read alignment, read pairing, and restriction cutting sites [21]. During Hi-CPro processing, reads were aligned to the Arabidopsis genome by using Bowtie 2 with the default parameters of Hi-CPro [22]. The restriction enzyme of each sample used for filtering restriction cutting sites is listed in Supplementary Table S3. To calculate the correlations between replicates and datasets, the matrices of 20 samples at 20-kb resolution were normalized using hicNormalize. High correlations were found among 20 Hi-C samples at 20-kb resolution, verify the stability of the large compartment of chromatin interactions across different tissues from previous studies (Supplementary Fig. S2) [23]. The normalized matrices were corrected using hicCorrectMatrix with the Knight–Ruiz balancing algorithm and default parameters. Both hicNormalize and hicCorrectMatrix are tools of HiCExplorer [24].

2.4. Identification of TAD and statistical analysis of TAD boundaries

The Hi-C matrices at 1-kb resolution were normalized and corrected using the tools and parameters mentioned in section 2.3. The corrected matrices at 1-kb resolution were used to define TADs by using hicFindTADs with the parameter “--correctForMultipleTesting fdr” [24]. The bed files of 1-kb TAD boundaries generated from hicFindTADs were further used to characterize the genomic features and the distribution of ChIP-seq-based depictions. To avoid statistical bias, the genomic background was created as a control. Two thousand 1-kb regions were randomly selected from non-TAD boundary regions with the same GC content as the TAD boundaries. This random selection was repeated 100 times. In the statistical test, TAD boundaries of 20 samples were compared with 100 non-TAD boundary region sets using a \( t \) test. To estimate TF binding peaks, histone mark deposition, and DHSs within the flanking 5 kb centered at the 5' end of the TAD boundary, TAD boundaries of 20 samples were merged according to their genomic coordinates. The overlapping tools and calculation of normalized values were the same as the estimation of ChIP-seq peaks within the flanking 5 kb of TSSs. DHSs were retrieved from PlantRegMap.
2.5. Identification of NR and HS genes

RNA-seq expression datasets of two heat-stress treatments (5-week-old plants subjected to 37 °C for 30 min and 30-day-old plants subjected to 38 °C for 6 h) were obtained from GSE85653 and GSE118298, respectively [26, 27]. The differentially expressed gene search function of EXPPath 2.0 was applied to identify HS genes [18]. For each expression dataset, HS genes were selected by using a t test to compare heat-stress treatment and a control sample with p ≤ 0.01 and fold change (log2) ≥ 1. The false discovery rate (FDR) was set to 0.1. A total of 668 HS genes were selected from two heat-stress treatments (Supplementary Table S4). The unpublished in-house dataset (22-day-old plants exposed to long-term temperature treatment at 23 °C, 28 °C, and 30 °C after seed stratification) was used to identify differentially expressed genes under long-term warm temperatures. By using a t test with p ≤ 0.01, fold change (log2) ≥ 1, and FDR = 0.5 on two replicates, 609 genes were defined as long-term warm-temperature-responsive genes (Supplementary Table S5). Additionally, microarray and RNA-seq data of 175 and 99 stress-related conditions, respectively, were retrieved from EXPPath 2.0 to identify NR genes [18]. The NR genes were filtered as follows: (1) genes with low expression (transcript per million < 1) in all stress-related conditions were discarded, and (2) selected genes were fold change (log2) ≤ 0.8 for all stress-related conditions in both microarray and RNA-seq data. The final number of NR genes was 148 (Supplementary Table S6).

3. Results

3.1. Differences in cis-regulatory elements between TFs and between protein-coding and non-coding genes

In our previous study, we constructed genome-wide landscapes of 53 individual TFs belonging to 16 TF families by systematically collecting ChIP-seq data with strict criteria and standard data processing (Supplementary Table S1) [15]. To characterize the genome-wide regulation of TFs, compositions of individual TF binding peaks were mapped onto 16 gene types. Approximately 34 % of the Arabidopsis genome was occupied by TF binding peaks, which were associated with 25,452 (92 %) protein-coding genes and 20,092 (48 %) non-coding genes. Among the 53 analyzed TFs, for 41 (76 %) TFs, their binding peaks were located at protein-coding genes (including their 1-kb flanking regions; Fig. 1A). Normally, on the basis of data from 89 % of TFs, small fractions of binding peaks were located at lncRNAs. Over 20 % of the binding peaks of AZF1, DELLA, AP1, FIE, and CCA1 were mapped to transposable elements (TEs; i.e., transposon fragments and transposable genes). The lncRNAs peaks demonstrated that BZIP28, FYH3, and SVP may play regulatory roles in rRNA (Supplementary Table S7). To further investigate whether TFs could reveal different binding patterns in protein-coding genes and non-coding genes, peak occurrences were estimated on the basis of subdivided genetic regions (Supplementary Fig. S1). With seven genetic regions of protein-coding genes, a general binding preference was observed in 70 % of TFs, widely existing in members of the bZIP, homeodomain, and NAM, NF-YB, and NF-YC families (Fig. 1B). This preference generally comprised approximately 40 % upstream 1 kb, 20 % downstream 1 kb, 10 % ≤ UTR, 5 % ≤ 3′ UTR, 9 % intergenic regions, and a small proportion of CDSs and introns. In contrast with most TFs, which use promoters as dominant regulatory regions, HB1, AZF1, HSFA1A, SVP, FIE, and TOC1 exhibited a major-
for their functions of gene silencing and heterochromatin condensation [35,41–43]. To better understand histone marks, the histone landscapes were mapped to the subdivided genetic regions of protein-coding and non-coding genes. Differing from the TFs which were largely located at upstream and downstream regions of protein-coding genes, the histone marks highly overlapped with CDSs and introns (Fig. 1B, 2B). Nevertheless, neither activating nor repressive histone marks exhibited consistent patterns in the genetic locations of protein-coding and non-coding genes (Fig. 2B, C). The depositions of seven histone marks (H2A.Z, H3K14ac, H3K23ac, H3K36ac, H3K36me3, H3K4me3, and H3K9ac) related to gene activation were located at the 5'0 UTR of protein-coding genes, in contrast with other histone marks (Fig. 2B). Moreover, the data revealed distinct histone combinations in each type of non-coding gene (Fig. 2D).

Fig. 1. The preferences of TF binding peaks. (A) The peak depositions of 16 gene types for 53 TFs. The TF families are marked in rectangular bars at the left of the row labels. The regions of protein-coding genes include their upstream, and downstream 1 kb. The percentages of binding peaks at genetic regions of protein-coding genes (B) and non-coding genes (C). For (A-C), genome annotation shows the genome coverage of 16 gene types in the Arabidopsis genome, including regions that are not annotated with any genes (intergenic regions). (D) The distributions of TF binding peaks within flanking 5 kb of TSSs (left) and TTSs (right) of protein-coding genes. The ‘‘5’’ and ‘‘5’’ of the x-axis stand for the sites at upstream 5 kb and downstream 5 kb from TSSs (or TTSs), respectively. The bin size is 100 bp. lncRNA, long non-coding RNA, ncRNA, non-coding RNA, rRNA, ribosomal RNA, snoRNA, small nuclear RNA, snRNA, small nucleolar RNA, tRNA, transfer RNA, TSS, transcription start site, TTS, transcription termination site.

C.-N. Chow, K.-C. Tseng, P.-F. Hou et al. Computational and Structural Biotechnology Journal 20 (2022) 4910–4920

4913
and H3K27me1, and activating H3.3 histone marks. Over half of the snoRNAs were packaged by activating H3K36ac, H3K36me3, H3K4me3, and H3K9ac. The histone combinations of antisense lncRNAs were more similar to protein-coding genes than were those of lncRNA, indicating that the coexpression of antisense lncRNAs and protein-coding genes may be caused by similar regulation of histone marks [44]. Overall, these findings indicated that the plants used different combinations of histone variants and modifications to wrap non-coding genes and that various types of non-coding genes may have distinct regulatory roles.

To further verify that TFs and histone marks could regulate transcription through different genetic regions of genes, the distributions of histone mark depositions within flanking 5 kb of TSSs (left) and TTSs (right) of protein-coding genes. The “-5” and “5” of the x-axis stand for the sites at upstream 5 kb and downstream 5 kb from TSSs (or TTSs), respectively. The bin size is 100 bp. IncRNA, long non-coding RNA. ncRNA, non-coding RNA. rRNA, ribosomal RNA. snRNA, small nuclear RNA. snoRNA, small nucleolar RNA. tRNA, transfer RNA. TSS, transcription start site. TTS, transcription termination site.

3.3. Bilateral symmetry of histone marks and enrichment of cis-regulatory elements on TAD boundaries

Hi-C sequencing revealed the high-ordered organization of chromatin in plants. Unlike mammalian cells demonstrating the coregulation of genes and the regulatory isolation of TADs, the
TAD boundaries in plant cells were enriched with activating genes and related to epigenetic regulation [45,46]. The TADs and interactions between promoters and enhancers allowed predicting the association between expression of genes and their regulation [13]. To characterize the genetic features of TAD boundaries, the TAD boundaries of *Arabidopsis* were identified from the public Hi-C sequencing samples from nine datasets (20 samples; Supplementary Table S3). Among the defined TAD boundaries at 1-kb resolution, over 35% of TAD boundaries were conserved in more than one Hi-C sample. The statistical analysis revealed that TAD boundaries were significantly more highly overlapped with protein-coding genes than were those of random regions (Fig. 3A). Notably, these TAD boundaries were significantly higher on upstream 1 kb, 5' UTR, and 3' UTR and significantly lower on CDSs and introns (Fig. 3B). These results indicated that TAD boundaries were potentially related to gene regulation through promoters and UTRs. The statistical analysis of non-coding genes revealed that TAD boundaries were abundant in antisense lncRNA, lncRNA, ncRNA, snRNA, snoRNA, and tRNA (Supplementary Fig. S10). However, the TEs were significantly located in the outer regions of TAD boundaries. This result indicated that the regulation of TEs might be less strict in TAD organization than in other non-coding genes, thus enabling TEs to change their positions. To further assess whether gene functions were different between genes inside and outside TAD bound-
aries, GO term enrichment analysis was applied. Unexpectedly, 14,627 genes within TAD boundaries were found to possess fundamental functions (Supplementary Table S9). By contrast, 12,818 genes outside TAD boundaries were particularly related to TF activity and responses to environmental stress (Supplementary Table S10).

To examine the associations between epigenetic regulation and TAD boundaries, 5-kb flanking regions from the 5’ end of TAD boundaries were mapped with the depositions of 19 histone marks. The results revealed the bilateral symmetry of 19 histone marks around the TAD boundaries (Fig. 3C). Most activating histone marks displayed broad enrichment of TAD boundaries, suggesting that these histone marks might be essential to maintaining gene activation within TAD boundaries. The remaining histone marks, especially the most repressive marks, resided at region 2 kb, far from TAD boundaries. However, the significant locations at upstream 1-kb and UTR regions raised the question of whether TAD boundaries were associated with TF binding sites. Hence, by mapping TAD boundaries with TF binding peaks, 66 % of TAD boundaries were observed to overlap with TF binding peaks by at least 1 bp. The binding peaks of 41 (77 %) TFs were found to perform the prominent enrichments inside TAD boundaries (Fig. 3D). Overall, the data illustrated the recruitment of epigenetic regulation and TF binding around TAD boundaries, suggesting the potential transcriptional activation function of TAD boundaries in Arabidopsis.

3.4. Substantial differences of the cis-regulatory and epigenetic regions for genes with distinct functions

In mammals and Drosophila, different properties (e.g., motif configurations and chromatin states) at promoter regions were found in genes with distinct functions, such as housekeeping genes, cell-specific genes, and developmental-related genes [1]. To characterize the regulatory regions of Arabidopsis genes with different functions, 668 HS genes were selected from two RNA-seq datasets (Supplementary Table S4) [26,27]. Because common plant housekeeping genes, such as ACT2 and TUB6, were differentially expressed under at least one condition in high-throughput data (Supplementary Table S11), 148 NR genes were identified from public RNA-seq and microarray expression datasets (Supplementary Table S6). Compared with NR genes, HS genes displayed longer DNA sequences on genes, exons, amino acid sequences, 5’ UTR, and 3’ UTR, as well as more exon numbers (Fig. 4A). By mapping TF binding peaks to TSSs and TTSs, TFs were observed to use distinct regions to regulate two gene groups (Fig. 4B, C). TFs prefer to regulate HS genes by using the regions centered on TSSs and TTSs whereas TFs control NR genes by locating at regions closer to or farther from 1-kb flanking regions, indicating that cis-regulatory elements around TTSs play key roles in regulating HS genes. The depositions of histone marks verified these differences (Fig. 4D, E). The activating H3K14ac, H3K23ac, H3K36ac, H3K36me3, H3K4me2, H3K4me3, H3K9ac, and repressive H3K27me1 were depleted in gene bodies of NR but highly enriched at both 5’ and 3’ ends of HS genes. Conversely, repressive H3K27me3 were preferentially enriched downstream of TSSs and upstream of TTSs of NR genes instead of HS genes. Histone variant H2A.Z, related to thermoresponse responses, exhibited similar enrichments at TSSs of both gene groups but was only enriched around TTSs of NR genes [47]. Overall, these results suggested that the sequences and chromatin states around TSSs and TTSs could indicate a separation between NR genes and HS genes. The enrichments of activating histone marks and depletions of repressive histone marks at both 5’ and 3’ ends of HS gene bodies may facilitate a quick response to environmental changes. Furthermore, the significant differences in gene structures and cis-regulatory regions were also observed between long-term warm-temperature-responsive genes and NR genes (Supplementary Figs. S11 and S12), which further verified the different properties of genes with distinct functions.

The differences between HS and NR genes were primarily based on ChIP-seq datasets sampled under normal conditions without any heat stimuli. Few TFs (i.e., HSFA1A, ABI5, SOC1, and ABF3) have been reported to regulate gene expressions under heat stress [48–53]. These results indicated that the differences in gene regulation between the two gene groups may happen not only under heat stress but also under normal conditions (Fig. 4B, C). To further investigate whether the regulatory regions change in response to heat stress, DHSs were used as open chromatin regions [12]. Under both heat stress and control, HS genes contained the enrichments of DHSSs at TSSs and TTSs (Supplementary Fig. S13), which was consistent with TF binding peaks (Fig. 4C). Compared with HS genes, DHSSs displayed scattered arrangements on the 5-kb flanking regions of NR genes. A similar phenomenon was observed in the distributions of TAD boundaries identified from replicates of heat stress and controls [13]. Although these experimental datasets (i.e., RNA-seq, ChIP-seq, Hi-C, and DHSSs) were generated through different experimental designs (temperatures, time as heat stimuli, and developmental stages of harvested plant tissues), all exhibited consistent differences between the NR and HS gene groups. Together, these results suggested that genes with distinct functions were substantially different in transcriptional regulatory regions, epigenetic regulations, and even TAD boundary organizations. This implied that, compared with NR genes, HS genes require more stringent regulatory regions and epigenetic environments to regulate their gene expressions, leading to accurate responses when plants encounter environmental changes. Moreover, these results indicated that the destiny of a gene is decided more by its sequence and chromatin states than by its environment.

4. Discussion

Generally, the promoters of protein-coding genes are the key regions to interact with TFs, and they play key roles in transcriptional regulation. However, whether the promoter is the key regulator for non-coding gene regulation remains unknown. On the basis of multiple experimental techniques and high-throughput sequencing, the genomic landscapes of TFs, histone marks, DHSs, and TAD boundaries were unveiled, enabling scientists to investigate the mechanisms underlying gene regulation in plants. In this study, we demonstrated the complexity of gene regulation through the genomic landscapes of TFs, histone marks, DHSs, and TAD boundaries. Fig. 5 integrates our results and illustrates our proposed general regulatory models of protein-coding and non-coding genes.

The results from the mapping the TF binding peaks on protein-coding genes suggested that most TFs bind to the upstream regions of TSSS, which are typical promoters, but in terms of the frequency of TF binding, upstream regions accounted for less than half of the binding peaks (Figs. 1 and 5A). The nonpromoter regions (i.e., 5’ UTRs, 3’ UTRs, downstream regions, CDSs, and introns), particularly the downstream regions of TTSs, were also used for TF binding. Moreover, activating histone marks were found to be highly enriched at the gene bodies (downstream of TSSS and upstream of TTSs) of protein-coding genes but not promoters. The different genetic region usages between TFs and histone marks may explain the failure in TF binding prediction through the chromatin states of binding sites [7]. Peak occupancies revealed that TFs could bind not only to protein-coding genes but also to non-coding genes. Unlike the promoters of protein-coding genes, TFs tended to bind exons of non-coding genes. In the proposed model, we used two different
gene types, antisense lncRNAs and transposable element genes, to illustrate the diversity of TF and epigenetic regulation in the non-coding genes (Fig. 5B). Overall, the regulation of protein-coding genes and non-coding genes suggested that the promoter region was not adequate for the construction of TF regulation and epigenetic regulation.

To further illustrate the complexity of gene regulation, we identified two gene groups with distinct gene expression patterns. NR and HS genes displayed differing usage of regulatory regions as TF binding and histone states (Fig. 5C). The construction of dynamic TF binding was a common method to develop the gene regulation of HS genes. Yet, ChIP-seq samples generated under both control and heat stress were lacking. To resolve this issue, we used DHSs defined under control and heat stress [12,54]. The results demonstrated that open chromatin regions were stable between control and heat-stress samples, illustrating that control of gene expression is dependent on both inherent cis-regulatory regions and environmental changes of trans-regulators.

As indicated by the identification criteria of TADs, the boundary of a TAD represents low-chromatin interaction between its right and left genomic regions [24]. Our results revealed that the TAD boundaries were enriched with several activating histone marks and TF binding and were depleted with repressive histone marks (Fig. 3B and 5D). This result may suggest that TAD boundaries pro-
vide the regions with a restricted number of chromatin interactions between adjacent regions and stable chromatin states to TF binding. Similar to DHSs, the TAD boundaries on NR genes exhibited more location flexibility than did HS genes, implying that the organization of chromatin interactions is also required for specific regions, such as promoters of TF binding.

The present study unveiled newly discovered information regarding high-throughput sequencing data and observed that no unique rules can perfectly explain the regulation of all regulators and all regulated genes. We believe that the integrated analysis of multiple factors will increase the understanding of gene regulation and aid in the predication of regulatory elements.

CRediT authorship contribution statement

Chi-Nga Chow: Conceptualization, Software, Formal analysis, Investigation, Writing – original draft, Visualization. Kuan-Chieh Tseng: Formal analysis, Investigation, Visualization. Ping-Fu Hou: Formal analysis, Investigation. Nai-Yun Wu: Software, For-
nal analysis. **Tzong-Yi Lee**: Conceptualization. **Wen-Chi Chang**: Conceptualization. Writing – review & editing.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgments**

We appreciate all the scientists whose work helped in the current study. We would also like to thank the Ministry of Science and Technology (MOST 108-2311-B-006-002-MY3 and MOST 111-2311-B-006-006) and National Cheng Kung University for financially supporting this research. We are also grateful to the National Center for High-Performance Computing for computer time and facilities.

**Author Contributions**

C.N.C. and W.C.C. designed the research; C.N.C. performed the research; C.N.C., K.C.T., N.Y.W., and P.F.H. analyzed and visualized the data; C.N.C. and W.C.C. wrote the paper; T.Y.L. and W.C.C. advised on the research.

**Appendix A. Supplementary data**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.08.058.

**References**

[1] Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Biol 2018;19(10):621–37.
[2] Sega P, Kruszka K, Szwec L, Szewykowska-Kulinska Z, Pacak A. Identification of transcription factors that bind to the S-UTR of the barley PHO2 gene. Plant Mol Biol 2020;102(1–2):73–88.
[3] Meng F, Zhao H, Zhu B, Zhang T, Yang M, Li Y, et al. Genomic editing of intronic RNA-binding protein ORC1 in rice, foxtail millet and maize suggests non-canonical function of plant chromatin domains. J Integr Plant Biol 2020;62(2):201–17.
[4] Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, et al. Galaxy Protoc Bioinformatics 2014;47:11–1–34.
[5] Tseng KC, Li GZ, Hung VC, Chow CN, Wu NY, Chien YY, et al. ExpPath 2.0: An Updated Database for Integrating High-Throughput Gene Expression Data with Biological Pathways. Plant Cell Physiol 2020;61(10):1818–27.
[6] Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30(1):207–10.
[7] Kodama Y, Shumway M, Leinonen R, C. International Nucleotide Sequence Database. In: Bork P, editor. PlantPAN3.0: a new database for plant transcription factors and regulatory elements. Plant Physiol 2011;157:247–52.
[8] Yin X, Romero-Campero FJ, de Los Reyes P, Yan P, Yang J, Tian G, Yang X, Mo X, et al. The histone code in Arabidopsis leaves under prolonged warming and heat shock. BMC Plant Biol 2020;20(1):86.
[9] Alahesh WS, Obomighie I, Blein T, Persad R, Cheremukhin I, Crespi M, et al. Arabidopsis HEAT SHOCK TRANSCRIPTION FACTOR1A1 regulates multiple developmental genes under benign and stress conditions. J Exp Bot 2018;69(11):2847–62.
[10] Kodaira KS, Qin F, Tran LS, Maruyama K, Kidokoro S, Fujita Y, et al. Arabidopsis Cys2/His2 zinc-finger proteins AZF1 and AZF2 negatively regulate abscisic acid-repressive and auxin-inducible genes under abiotic stress conditions. Plant Physiol 2011;155:474–52.
[11] Fan M, Bai MY, Kim JG, Wang T, Oh E, Chen L, et al. The bHLH transcription factor HB1 mediates the trade-off between growth and pathway-associated molecular pattern-triggered immunity in Arabidopsis. Plant Cell 2014;26(2):828–41.
[12] Liu J, Feng L, Li J, He Z. Genetic and epigenetic control of plant heat responses. Front Plant Sci 2015;6:267.
[13] Liu Y, Ma M, Li G, Yuan X, Xie Y, Wei H, et al. Transcription factors PHY3 and FAR1 regulates light-induced CIRCADIAN CLOCK ASSOCIATED1 gene expression in Arabidopsis. Plant Cell Physiol 2011;52(5):1464–78.
[14] Deng W, Buzas DM, Ying H, Robertson M, Taylor J, Peacock WJ, Dennis ES, Hessellie C. Arabidopsis polycomb repressive complex 2 binding sites contain putative GAGA factor binding motifs within coding regions of genes. BMC Genomics 2013;14:593.
[15] Klemm SL, Shipony Z, Greenleaf JW. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet 2019;20(4):207–20.
[16] Kim YJ, Wang R, Guo L, Li X, Yu D, Wang H, et al. POWERDRESS and HDA9 interact and promote histone H3 deacetylation at specific genomic sites in Arabidopsis. Proc Natl Acad Sci U S A 2016;113(51):14858–63.
[17] Yelagandula R, Stroud H, Lescot S, Zhou K, Feng S, Zhong X, et al. The histone variant H2A.Z defines heterochromatin and promotes chromatin condensation in Arabidopsis. Cell 2014;158(1):98–109.
[18] Ha M, Ng DW, Li WH, Chen CJ. Coordinated histone modifications are associated with gene expression variation within and between species. Genome Res 2011;21(4):596–8.
[19] Mahrez W, Arelanno MS, Moreno-Romero J, Nakamura M, Shu H, Nannini P, et al. H3K36ac is an evolutionary conserved plant histone modification that marks group III genes. Plant Biol 2015;17(3):1566–77.
[20] Zhang F, Qi B, Wang L, Zhao B, Rode S, Rigandin ND, Eckler JR, Qiao H. EIN2-dependent regulation of acetylation of histone H3K14 and non-canonical histone H3K23 in ethylene signalling. Nat Commun 2016;7:13018.
[21] Soria G, Polo SE. Almouzni G. Prime, repair, restore: the active role of chromatin in the DNA damage response. Mol Cell 2014;55(4):324–34.
[22] Yin X, Romero-Campero FJ, de Los Reyes P, Yan P, Yang J, Tian G, Yang X, Mo X, Zhao S, Calonje M, Zhou Y. H2A.K121ub in Arabidopsis associates with a less accessible chromatin state at transcriptional regulation hotspots. Nat Commun 2021;12(1):315.
[23] Karker B, Dhilon T, Colas I, Cook N, Milne I, Milne L, et al. Chromatin state analysis of the barley epigenome reveals a higher-order structure defined by H3K27me3 and H3K27me2. Plant J 2015;84(1):111–24.
[24] Stroud H, Otero S, Desvoyes B, Ramirez-Parra E, Jacobsen SE, Gutierrez C. Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. Proc Natl Acad Sci U S A 2012;109(14):5370–5.
Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y. Global identification of Arabidopsis lincRNAs reveals the regulation of MAP4 by a natural antisense RNA. Nat Commun 2018;9(1):5056.

Dong P, Tu X, Liang Z, Kang BH, Zhong S. Plant and animal chromatin three-dimensional organization: similar structures but different functions. J Exp Bot 2020;71(17):5119–28.

Wang C, Liu C, Roqueiro D, Grimm D, Schwab R, Becker C, et al. Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. Genome Res 2015;25(2):246–56.

Kim JM, Sasaki T, Ueda M, Sako K, Seki M. Chromatin changes in response to drought, salinity, heat, and cold stresses in plants. Front Plant Sci 2015;6:114.

Kim JB, Kang JY, Kim SY. Over-expression of a transcription factor regulating ABA-responsive gene expression confers multiple stress tolerance. Plant Biotechnol J 2004;2(5):459–66.

Ohama N, Sato H, Shinozaki K, Yamaguchi-Shinozaki K. Transcriptional regulatory network of plant heat stress response. Trends Plant Sci 2017;22(1):53–65.

Sakuma Y, Maruyama K, Qin F, Osakabe Y, Shinozaki K, Yamaguchi-Shinozaki K. Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. Proc Natl Acad Sci USA 2006;103(49):18822–7.

Skubacz A, Daszkowska-Golec A, Szarejko I. The role and regulation of ABI5 (ABA-insensitive 5) in plant development, abiotic stress responses and phytohormone crosstalk. Front Plant Sci 2016;7:1884.

Liu HC, Liao HT, Charing Y. The role of class A1 heat shock factors (HSFA1s) in response to heat and other stresses in Arabidopsis. Plant Cell Environ 2011;34(5):738–51.

Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 2017;45(D1):D1040–5.