

# Determinants of nucleosome positioning and their influence on plant gene expression

Ming-Jung Liu,<sup>1</sup> Alexander E. Seddon,<sup>1</sup> Zing Tsung-Yeh Tsai,<sup>1</sup> Ian T. Major,<sup>2</sup> Monique Floer,<sup>3</sup> Gregg A. Howe,<sup>2,3</sup> and Shin-Han Shiu<sup>1</sup>

<sup>1</sup>Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824, USA; <sup>2</sup>DOE-Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824, USA; <sup>3</sup>Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA

Nucleosome positioning influences the access of transcription factors (TFs) to their binding sites and gene expression. Studies in plant, animal, and fungal models demonstrate similar nucleosome positioning patterns along genes and correlations between occupancy and expression. However, the relationships among nucleosome positioning, *cis*-regulatory element accessibility, and gene expression in plants remain undefined. Here we showed that plant nucleosome depletion occurs on specific 6-mer motifs and this sequence-specific nucleosome depletion is predictive of expression levels. Nucleosome-depleted regions in *Arabidopsis thaliana* tend to have higher G/C content, unlike yeast, and are centered on specific G/C-rich 6-mers, suggesting that intrinsic sequence properties, such as G/C content, cannot fully explain plant nucleosome positioning. These 6-mer motif sites showed higher DNase I hypersensitivity and are flanked by strongly phased nucleosomes, consistent with known TF binding sites. Intriguingly, this 6-mer-specific nucleosome depletion pattern occurs not only in promoter but also in genic regions and is significantly correlated with higher gene expression level, a phenomenon also found in rice but not in yeast. Among the 6-mer motifs enriched in genes responsive to treatment with the defense hormone jasmonate, there are no significant changes in nucleosome occupancy, suggesting that these sites are potentially preconditioned to enable rapid response without changing chromatin state significantly. Our study provides a global assessment of the joint contribution of nucleosome occupancy and motif sequences that are likely *cis*-elements to the control of gene expression in plants. Our findings pave the way for further understanding the impact of chromatin state on plant transcriptional regulatory circuits.

[Supplemental material is available for this article.]

The dynamic and specific binding of transcription factors (TFs) to DNA sequences plays a vital role in the temporal and spatial regulation of gene expression. Because DNA is wrapped in nucleosomes (Richmond and Davey 2003), the modulation of nucleosome occupancy influences genome accessibility, thus creating a direct means to control gene expression. Consistent with the role of nucleosomes in regulating gene expression, global nucleosome mapping studies in yeast and humans have shown that lower nucleosome occupancy in promoter regions is correlated with elevated gene expression (Lee et al. 2007; Valouev et al. 2011). For example, loss of yeast histone proteins in promoter regions leads to elevated expression of approximately 2500 genes (Gossett and Lieb 2012), demonstrating the global effect of nucleosomes on gene expression. In addition, highly expressed genes tend to have phased nucleosomes along the gene body and a ~150-bp nucleosome-depleted region (NDR) immediately upstream of the transcription start site (TSS). This nucleosome occupancy pattern reflects the competition between nucleosome formation and RNA polymerase or general TFs binding to core promoter regions (Valouev et al. 2011; Struhl and Segal 2013).

There are at least two factors determining nucleosome occupancy in the genome. The first is intrinsic sequence property. In yeast and human studies of the *in vitro* reconstruction of genomic DNA and histones into nucleosomes, poly(dA/dT) stretches repel

and G/C-rich sequences promote nucleosome formation (Kaplan et al. 2009; Valouev et al. 2011; Zhang et al. 2011; Hughes et al. 2012; Struhl and Segal 2013). The second is cellular *trans*-acting factors including transcription factors and chromatin remodelers that also affect nucleosome occupancy, overriding the influence of intrinsic sequence properties. For example, binding sites for human TFs tend to occur in nucleosome-depleted and G/C-rich regions, in contrast to findings based on *in vitro* reconstitution experiments (Valouev et al. 2011; Wang et al. 2012). The competition between nucleosome formation and TF binding to G/C-rich regions provides a mechanism to regulate transcriptional activity (Wang et al. 2012). This is illustrated by the finding that insertion or deletion of the nucleosome-disfavoring sequence poly(dA/dT) around TF binding sites (TFBSs) significantly affects transcription in yeast (Raveh-Sadka et al. 2012), showing the importance of chromatin structure as a determinant of gene expression.

In plants, genome-wide nucleosome positions have been determined for *Arabidopsis thaliana* (Chodavarapu et al. 2010; Li et al. 2014), rice (*Oryza sativa*) (Wu et al. 2014), and maize (*Zea mays*) (Fincher et al. 2013; Vera et al. 2014). These studies examined the relationship between nucleosome positions and DNA methylation, known TFBSs, and gene expression (Chodavarapu et al.

**Corresponding author:** shius@msu.edu

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.188680.114>.

© 2015 Liu et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

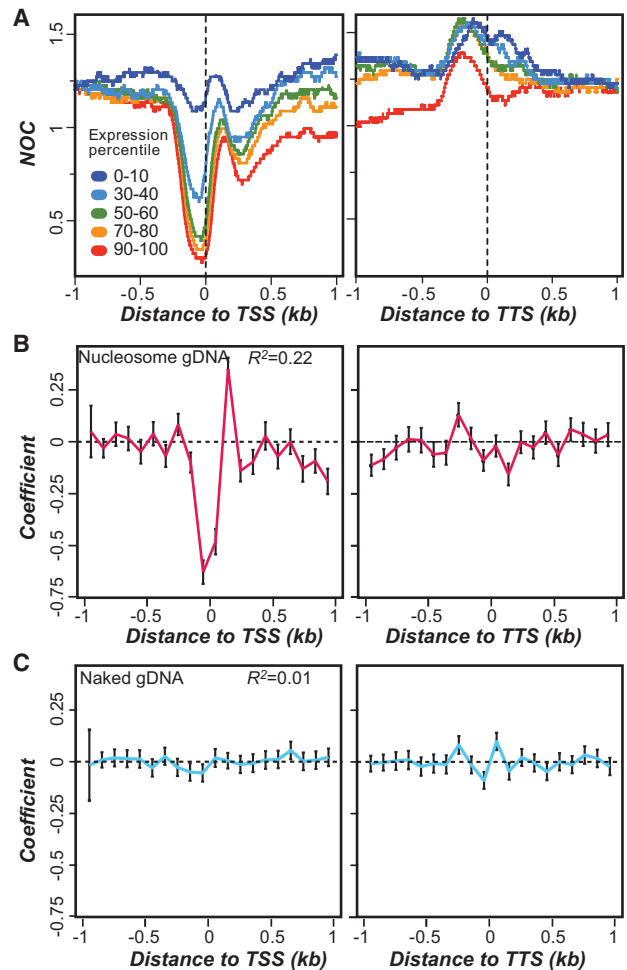
2010; Fincher et al. 2013; Li et al. 2014; Vera et al. 2014; Wu et al. 2014). In particular, it was shown that the nucleosome density around TSSs and within gene bodies correlates with gene expression in *A. thaliana* and maize (Li et al. 2014; Vera et al. 2014). It has also been shown that phased nucleosomes flank known TFBSs and DNase I hypersensitive sites (DHSs) in rice, which are likely bound by regulatory proteins (Wu et al. 2014). In addition, studies of DHSs in *A. thaliana* have revealed the relationships between chromatin accessibility and environmentally sensitive or tissue-specific DNA elements (Zhang et al. 2012; Sullivan et al. 2014). Nonetheless, there are several questions relevant to the relationships between nucleosome occupancy, *cis*-regulatory sequences, and gene expression that are yet to be addressed in plants or other model systems. The first is to what extent knowledge of nucleosome occupancy may explain gene expression in a statistical context. Second, although TFBSs were found to be surrounded by phased nucleosomes in plants (Wu et al. 2014), it remains unclear if there are additional sequence motifs that tend to be located in nucleosome-depleted regions, how these motifs affect gene expression, and whether the relationships between sequence motifs and nucleosome depletion are influenced by where the motifs are located. Third, the extent to which nucleosome occupancy in plants is influenced by environmental stimuli remains to be determined.

To address the preceding questions, we profiled genome-wide nucleosome positioning and transcript abundances in *A. thaliana* to investigate the relationships between nucleosome occupancy, 6-mer motif sequences resembling TF binding sites, and gene expression.

## Results and discussion

### Nucleosome occupancy and gene expression

To assess nucleosome positions, micrococcal nuclease (MNase) was used to digest nucleosome-bound genomic DNA (gDNA) isolated from *A. thaliana* rosette leaves, and the resulting DNA fragments were sequenced (“nucleosome gDNA” sample). For comparison, a “naked gDNA” sample stripped of proteins was also digested with MNase and sequenced (see Methods). Consistent with the expected size of mononucleosomes (Richmond and Davey 2003), the most prominent band from nucleosome gDNA samples was ~150 bp (Supplemental Fig. S1A), in contrast to the naked gDNA samples (Supplemental Fig. S1B). The nucleosome gDNA read coverage per base—referred to as the Nucleosome Occupancy (NOC) score (see Methods)—revealed NDRs immediately upstream of TSS and a prominent +1 peak (Supplemental Fig. S1C,D), consistent with earlier studies (Yuan et al. 2005; Valouev et al. 2011; Fincher et al. 2013; Li et al. 2014; Wu et al. 2014). In contrast, the naked gDNA sample did not show these patterns (Supplemental Fig. S1C,D). To evaluate the relationship between nucleosome occupancy and gene expression, genes were binned according to transcript levels. Compared to genes with lower transcript abundance, genes with higher transcript abundance tend to have lower nucleosome occupancy in NDRs and genic regions, a well-positioned nucleosome (larger differential between the peak and surrounding trough) immediately downstream from the TSS, and a characteristic occupancy pattern surrounding the transcription termination site (TTS) (Fig. 1A), which is consistent with previous studies in *A. thaliana*, rice, maize, and human (Valouev et al. 2011; Fincher et al. 2013; Li et al. 2014; Wu et al. 2014). Taken together, these findings indicate our nucleosome data are of comparable quality to published studies.



**Figure 1.** Relationships between nucleosome occupancy and gene expression. (A) Nucleosome occupancy profiles in regions flanking the transcription start site (TSS, left) and transcription termination site (TTS, right) of protein coding genes with different transcript levels (expression percentile in which the 0–10 category includes genes with the lowest expression level): (NOC) median nucleosome occupancy score of genes per base. (B) Regression coefficient of a linear regression model for gene expression prediction using nucleosome occupancy in 100-bp bins flanking TSS (left) and TTS (right) as predictors: ( $R^2$ ) coefficient of determination of the overall model; (error bars) the 5th and 95th confidence interval. (C) Same as B, except that naked gDNA data were used.

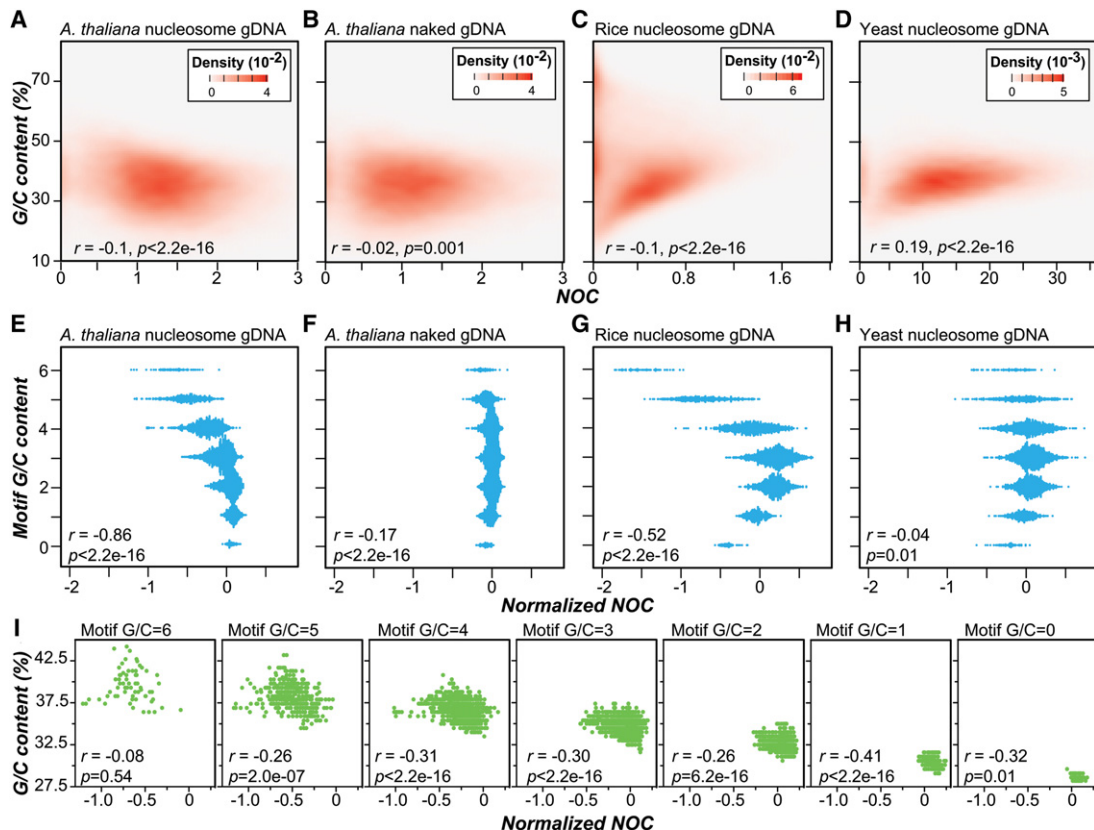
Although it is clear that nucleosome occupancy is negatively correlated with gene expression, it is not known to what extent nucleosome occupancy differences among genes may explain variation in gene expression in a quantitative model. To this end, a linear regression model was established by examining nucleosome occupancy in 2-kb regions surrounding both TSS and TTS (Fig. 1B). The coefficient of determination ( $R^2$ ) of the regression model is 0.22 (Fig. 1B), indicating that 22% of the transcript level variance can be explained by nucleosome occupancy (compared to  $R^2 = 0.01$  for the model based on naked gDNA data) (Fig. 1C). Note that in the regression model, the gene expression level is log transformed, indicating its nonlinear relationship to nucleosome occupancy (see Methods). In addition, the sequence bins differ greatly in how they contribute to the regression model. For example, the nucleosome occupancy in regions 100 bp upstream of and downstream from the TSS was significantly negatively correlated with

expression level, whereas the occupancy in the region corresponding to the +1 peak was strongly positively correlated compared to surrounding regions (Fig. 1B). Interestingly, correlation between nucleosome occupancy in genic region bins and expression level tends to be weaker but significantly negatively correlated, indicating that the extent of nucleosome occupancy in genic regions may also impact gene expression (Fig. 1B). Our results reveal that differences in nucleosome occupancy alone can explain expression-level differences to some extent, but ~78% of the expression level variance remains unaccounted for. Considering that chromatin accessibility influences how TFs interact with *cis*-regulatory sequences (Raveh-Sadka et al. 2012), this raises a question whether nucleosome occupancy on specific sequence motifs within the promoter could better explain gene expression among other contributing factors, such as RNA stability. To further reveal the relationship among nucleosome occupancy, specific sequences, and gene expression, we next determined how sequence context affects nucleosome occupancy.

### G/C content and nucleosome occupancy

The nucleosome occupancy in yeast and human genomes correlates with intrinsic sequence properties where G/C-rich sequences tend to have a higher nucleosome occupancy compared to

A/T-rich sequences (Mavrich et al. 2008; Valouev et al. 2011; Wang et al. 2012). Surprisingly, the nucleosome occupancy among randomly chosen 150-bp genomic regions in *A. thaliana* is significantly negatively correlated with G/C content (Pearson's correlation coefficient,  $r = -0.1$ ,  $P < 2.2 \times 10^{-16}$ ) (Fig. 2A). For comparison, this pattern is much weaker in naked gDNA ( $r = -0.02$ ,  $P = 1 \times 10^{-3}$ ) (Fig. 2B). To address whether this is a general feature in plants, we examined nucleosome occupancy from rice (Wu et al. 2014) that diverged from the *A. thaliana* lineage ~200 million years ago (Wolfe et al. 1989). Similar to *A. thaliana*, the rice nucleosome occupancy is negatively correlated with G/C content ( $r = -0.1$ ,  $P < 2.2 \times 10^{-16}$ ) (Fig. 2C). However, among rice genomic regions with 30%–40% G/C, there is a positive correlation ( $r = 0.21$ ,  $P < 2.2 \times 10^{-16}$ ), a pattern that is also found in human (Wang et al. 2012). The correlation coefficients ( $r^2$ ) in *A. thaliana* and rice are both about 0.01, showing that G/C content explains only ~1% of nucleosome occupancy differences (Fig. 2A,C). Nonetheless, the overall negative correlation between G/C content and nucleosome occupancy in plants is in contrast to the positive correlation seen in budding yeast ( $r = 0.19$ ,  $P < 2.2 \times 10^{-16}$ ) (Fig. 2D), suggesting that intrinsic sequence properties may have a weaker influence on plant nucleosome positioning compared to fungi. Given the small effect size, the technical difference among studies may contribute to some of the differences between plant and fungi.



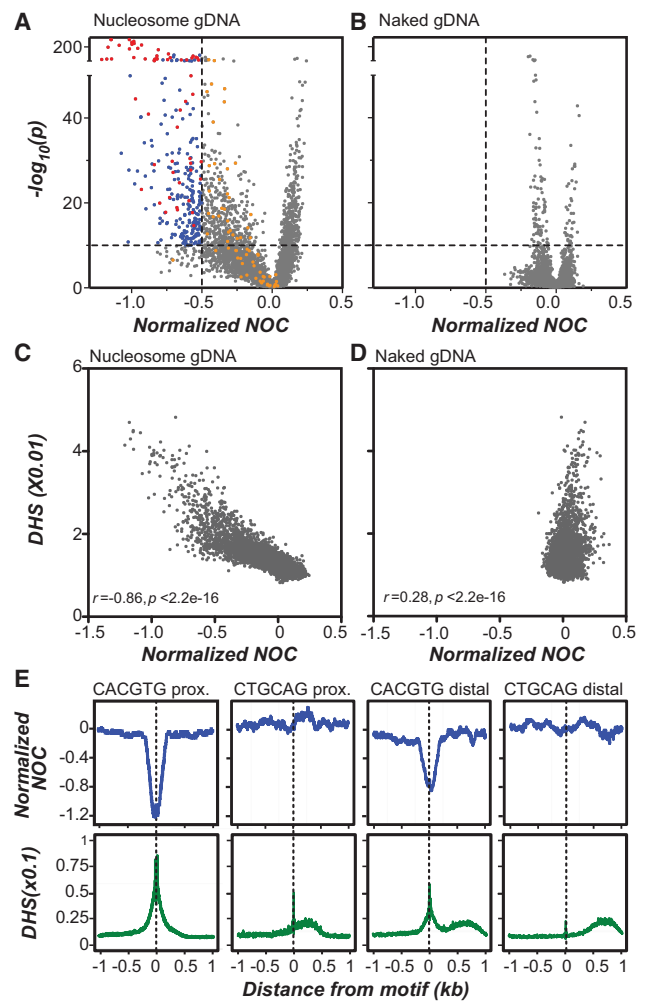
**Figure 2.** The correlation between nucleosome occupancy and G/C content in *A. thaliana*, rice, and yeast. (A–D) The correlation between G/C content and nucleosome occupancy determined with *A. thaliana* nucleosome gDNA (A), *A. thaliana* naked gDNA (B), rice nucleosome gDNA (C), and yeast nucleosome gDNA data (D). The density plots indicate the probability density of values from 36,000 randomly chosen 150-bp genomic regions. (E–H) Relationships between the normalized NOC scores (NOCs) and the numbers of G/Cs in 6-mers determined with the data indicated in A–D. The normalized NOC score is controlled for background nucleosome occupancy (see Methods). (I) Relationships between the percent G/C content of 200-bp regions centered on 6-mer motif sites and the medians of normalized NOCs of 6-mer motifs with different numbers of G/C nucleotides. ( $r$ ) Pearson's correlation coefficient.

Regions bound by regulatory factors in human tend to have high G/C content and low nucleosome occupancy (Valouev et al. 2011; Wang et al. 2012). In addition, nucleosome occupancy is influenced by chromatin remodelers and transcription factors (Valouev et al. 2011; Wang et al. 2012; Struhl and Segal 2013). Our findings (Fig. 2A–D) suggest that binding of *trans*-acting factors to specific sequence motifs in high G/C content regions is an important determinant of nucleosome positioning in *A. thaliana*.

To determine whether the extent of nucleosome occupancy in *A. thaliana* is correlated with specific sequence motifs, which would suggest the involvement of DNA binding *trans*-acting factors, we computed a normalized NOC score (see Methods) representing the degree of nucleosome occupancy in 200-bp regions centered on each possible 6-mer sequence after adjustment for the background occupancy (Supplemental Fig. S2). This normalization is necessary because NOC of a sequence is influenced by the region in which it is located (e.g., the NOCs in NDRs are lower in general) (Fig. 1A). Similar to the negative correlation between G/C content and nucleosome occupancy in randomly chosen genomic regions (Fig. 2A), there was a significantly stronger negative correlation between the nucleosome occupancy and number of G/C nucleotides in the 6-mer motifs in *A. thaliana* ( $r = -0.86$ ,  $P < 2.2 \times 10^{-16}$ ) (Fig. 2E) and rice ( $r = -0.52$ ,  $P < 2.2 \times 10^{-16}$ ) (Fig. 2G). In contrast, the correlations are substantially weaker in *A. thaliana* naked gDNA ( $r = -0.17$ ,  $P < 2.2 \times 10^{-16}$ ) (Fig. 2F) and the yeast nucleosome gDNA ( $r = -0.04$ ,  $P = 0.01$ ) (Fig. 2H). We should emphasize that, among 6-mer motifs with the same G/C content, some 6-mers have higher degrees of nucleosome depletion than other motifs (Fig. 2E), suggesting that specific sequence motifs influence the nucleosome occupancy pattern. In addition, the negative correlation between 6-mer motif G/C content and normalized NOC within motif-centered regions is not simply due to the G/C content of regions flanking the 6-mer sites, because G/C content of the motif-centered regions can only explain 0.64% (6-mer G/C = 6) to 6.8% (6-mer G/C = 2) of the variation of normalized NOCs (Fig. 2I). The nucleosome position patterns associated with motifs are not due to enzyme digestion or sequencing bias because the pattern is not observed in the naked gDNA data set (Fig. 2B,F). Taken together, our results suggest that plants are likely distinct from human and yeast in the relative influence of intrinsic and *trans*-acting factors on nucleosome positioning. In addition, specific-sequence motifs, particularly those with high G/C content, tend to be associated with nucleosome-depleted regions in plants.

### Relationship between 6-mer nucleosome occupancy and DNase I hypersensitivity

The binding of TFs to DNA influences nucleosome positioning and shapes the chromatin architecture around DNA regulatory elements (Struhl and Segal 2013; Hughes and Rando 2014). In mice and humans, although TFBSs tend to be nucleosome depleted, binding sites of some TFs tend to be nucleosome enriched (Teif et al. 2012; Wang et al. 2012). To identify *A. thaliana* 6-mer motifs that are significantly enriched in nucleosome-depleted or nucleosome-enriched regions, we compared the NOCs of 200-bp regions centered on a 6-mer motif to those of regions of randomly chosen genes with the Mann-Whitney *U* test (Fig. 3A; Supplemental Fig. S2; Methods). We identified 272 6-mer motifs located in the proximal promoter regions (TSS to 500 bp upstream) that have significantly lower nucleosome occupancy compared to regions



**Figure 3.** Degrees of nucleosome depletion around 6-mer motif sites and the relationships between occupancy and DNase I hypersensitivity (DHS) in *A. thaliana*. (A) Volcano plot showing the relationships between normalized NOC scores (*x*-axis) and the statistical significance ( $-\log_{10}$  [*P*-value], *y*-axis) testing whether the distribution of normalized NOCs of 200-bp regions around sites of each 6-mer differs from the normalized NOC distribution of randomly chosen sites with the same relative positions (Mann-Whitney *U* test). The NOC score here is determined according to the nucleosome gDNA data. (Red) 6-mers that have significant nucleosome depletion ( $P < 1 \times 10^{-10}$ ), have normalized NOC scores  $< -0.5$ , and are enriched among the top 10% highly expressed genes (Type I,  $n = 59$ ); (blue) 6-mers with significant nucleosome depletion but not enriched among highly expressed genes (Type II,  $n = 213$ ); (orange) 6-mers enriched among the highly expressed genes but with no significant nucleosome depletion (Type III,  $n = 66$ ). (B) Volcano plot as in A but determined using naked gDNA data. (C) Relationships between normalized NOC and DHS scores for regions surrounding 6-mer motifs based on nucleosome gDNA data. (D) Same as in C except naked gDNA data were used. (E) The normalized NOC and DHS (per base) profiles in regions surrounding genomic sites of the G-box (CACGTG) and a shuffled G-box sequence (CTGCAG) located in proximal (500 bp upstream of TSS) and distal (1 kb to 500 bp upstream of TSS) promoters. The *x*-axis represents the distance to motif sites.

neighboring randomly chosen genes (Mann-Whitney *U* test,  $P < 1 \times 10^{-10}$  and normalized NOC  $< -0.5$ ) (Fig. 3A, upper left quadrant). These 272 6-mer motifs are defined as significantly nucleosome-depleted motifs. On the other hand, we did not identify a single 6-mer to be enriched in regions with significantly higher nucleosome



occupancy than background (Fig. 3A). For comparison, no 6-mer motifs are significantly depleted or enriched in the naked gDNA sample based on the same criteria (Fig. 3B). Importantly, using a DHS data set from *A. thaliana* leaf samples (Zhang et al. 2012), we found a strong negative correlation between DHS and normalized NOC in nucleosome gDNA samples ( $r = -0.86$ ,  $P < 2.2 \times 10^{-16}$ ) (Fig. 3C; Supplemental Fig. S3B), but not in naked gDNA ( $r = 0.28$ ,  $P < 2.2 \times 10^{-16}$ ) (Fig. 3D). DHS is considered to be a hallmark of regulatory DNA elements (Gross and Garrard 1988). In *A. thaliana* and rice, the DHSs contain the *cis*-regulatory elements bound by known TFs (Zhang et al. 2012; Sullivan et al. 2014; Wu et al. 2014). Despite the differences in digestion mechanism and cleavage bias between MNase for determining nucleosome occupancy (Chung et al. 2010; Vierstra et al. 2014) and DNase I for revealing open chromatic regions (Sung et al. 2014; Vierstra et al. 2014), the positive correlation between motif-specific nucleosome depletion and heightened chromatin accessibility supports the notion that nucleosome-depleted motifs may be bound by TFs.

We next used the G-box motif (CACGTG) as an example to determine whether the nucleosome depletion patterns are specific for some sequence motifs and its relationship with DHS. G-box is recognized by bZIP and basic helix-loop-helix TFs including MYC2 (Jakoby et al. 2002; Toledo-Ortiz et al. 2003; Godoy et al. 2011). The normalized NOC in regions surrounding the G-box was clearly reduced and was associated with a correspondingly higher DHS (Fig. 3E). In contrast, regions surrounding the CTGCAG motif, which has a G/C content identical to that of the G-box, are not nucleosome depleted (Fig. 3E). To determine if similar nucleosome depletion and DHS patterns occur for motifs in regions with no NDRs, we defined the regions between 1.0 and 0.5 kb upstream of TSS as the “distal promoter.” Interestingly, nucleosome occupancy and DHS profiles of the CACGTG motif found in distal promoter regions are similar to the proximal promoter regions, although the signal is weaker (Fig. 3E). We also observed similar trends of nucleosome occupancy and DHS profiles between proximal and distal promoter regions for the GGGCCG motif (lowest normalized NOC = -1.2) and the TCTATA motif (highest normalized NOC = 0.2) (Supplemental Fig. S3A). These findings suggest that some motif sequences are specifically associated with nucleosome depletion regardless of their distance from the TSS, possibly due to binding of *trans*-acting factors as suggested by their correlations to DHS.

### Nucleosome depletion and 6-mer motif location

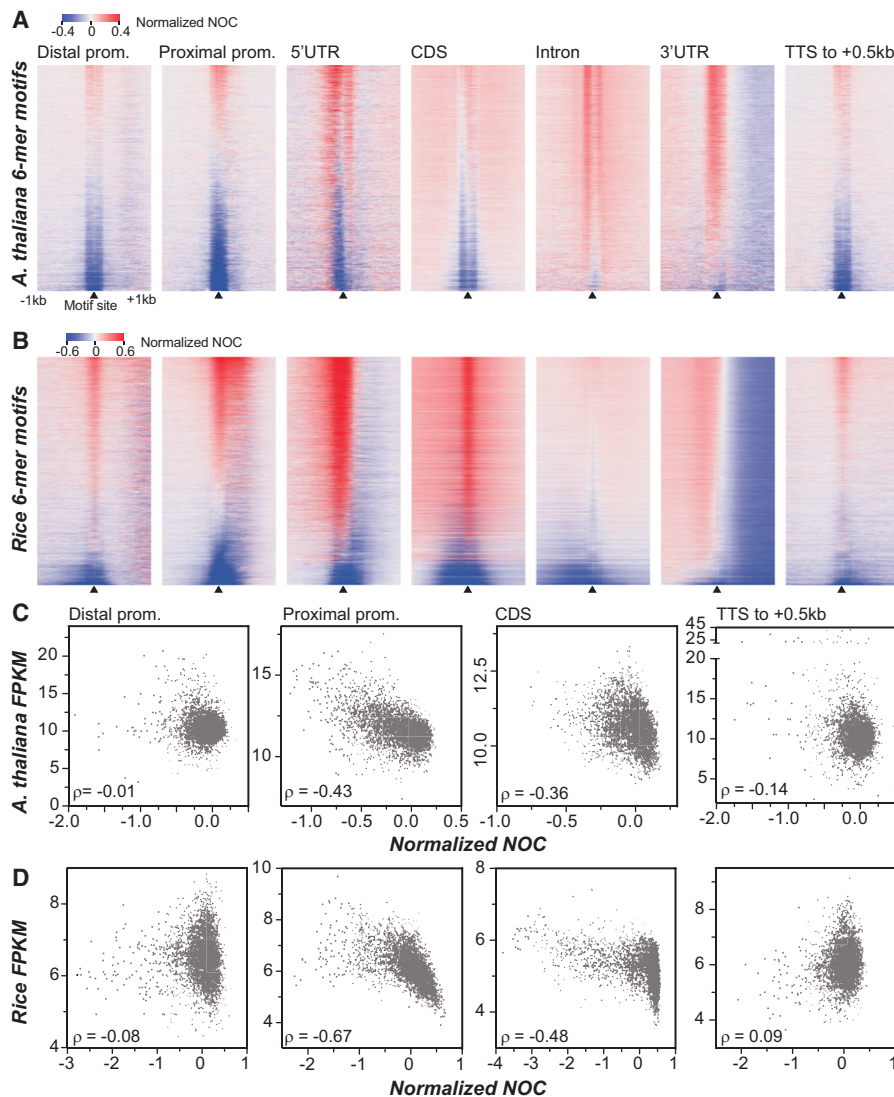
We showed that the nucleosome depletion patterns associated with the G-box and other motif sequences are similar between proximal and distal promoter regions (Fig. 3E; Supplemental Fig. S3A). In addition, *A. thaliana* DHSs are located not only in proximal promoter regions but also within the gene body or in intergenic regions more distant from genes (Zhang et al. 2012; Sullivan et al. 2014). Furthermore, TFBSs located in gene bodies are important for transcriptional regulation (Stergachis et al. 2013). To determine whether the nucleosome depletion patterns specific to 6-mer motifs in proximal promoter regions are generally found in other locations, we calculated normalized NOC for each base in 2-kb regions surrounding each 6-mer motif. Although the degree of nucleosome depletion among sites is the most pronounced in proximal promoters, strong nucleosome depletion patterns are also evident in distal promoters, coding sequences (CDS), and TTS 3' but are not apparent in UTRs and introns (Fig. 4A). Note that in 5' UTR, we observed a depletion pattern immedi-

ately before the motif sites, mostly due to the NDRs (Supplemental Fig. S4). These results are consistent with the finding that TSS, CDS, and intergenic regions have the highest proportion of DHS footprints (Sullivan et al. 2014).

Our findings highlight the potential contribution of *cis*-regulatory motifs not only in promoter regions but also in the gene body and regions downstream from TTS to the transcriptional regulation of *A. thaliana* genes. In addition, we showed that the nucleosome depletion pattern is motif specific in the promoter regions (Fig. 3). These two observations raise the question whether the nucleosome depletion associated with specific motif sites in promoters also occurred globally regardless of the genomic regions in which the motifs were located. To assess this, we compared the normalized NOC scores of a 6-mer motif located among different genomic regions. We found that a motif with a stronger depletion pattern in proximal promoter regions tends to have lower overall nucleosome occupancy in other regions also (Spearman's rank correlation coefficient,  $\rho = 0.67$ – $0.88$ , between proximal and all other region pairs,  $P < 2.2 \times 10^{-16}$ ) (Supplemental Table S1). Thus, although different regions have very distinct background nucleosome occupancies, the presence of certain motifs predicts, at least in part, the extent of nucleosome occupancy regardless of where the 6-mer motif is located (Fig. 4A; Supplemental Table S1). To further explore whether this pattern of motif-specific nucleosome depletion is a general feature of other eukaryotes, we analyzed MNase-seq data sets from rice (Wu et al. 2014) and budding yeast (Gossett and Lieb 2012). In rice, the depletion patterns in different regions are highly similar to those in *A. thaliana* (Fig. 4A,B). In addition, a 6-mer motif associated with low occupancy in *A. thaliana* tends to have a low NOC score in rice (between region pairs across species,  $\rho = 0.21$ – $0.71$ ,  $P < 2.2 \times 10^{-16}$ ) (Supplemental Table S1). In contrast, the motif-specific nucleosome depletion pattern between *A. thaliana* and budding yeast is only weakly negatively correlated ( $\rho = -0.08$ – $0.12$ ,  $P < 4.8 \times 10^{-7}$ ) (Supplemental Table S1). This would suggest that similar motifs are targeted by *trans* factors in plants but not in yeast.

### Motif-specific nucleosome depletion and gene expression

Our findings thus far indicate that the phenomenon of motif-specific nucleosome depletion is likely common among land plant species and fungi but different sets of sequences are involved. In addition, this phenomenon is particularly pronounced when motif sequences are located in promoters, CDS, and TTS 3', which is highly similar to the pattern observed in earlier studies based on analysis of DHS distributions (Sullivan et al. 2014). This raises the possibility that motifs located in nucleosome-depleted regions are *cis*-regulatory elements involved in regulating gene expression. To assess this possibility, we calculated for each 6-mer motif the median expression level of genes containing the motif in a particular region of the gene and the normalized NOC of the motif for the same region. Consistent with our interpretation that motif-specific nucleosome depletion patterns are relevant to gene expression regulation, in *A. thaliana* there is a significant correlation between median expression level and normalized NOC in the proximal promoter and CDS (Fig. 4C). The correlation between expression and motif-specific nucleosome depletion in proximal promoter and CDS is even stronger in rice (Fig. 4D). However, this is not the case in the distal promoter or downstream TTS regions (Fig. 4C,D). This is similar to findings in human showing that the number of DHS footprints within CDS is correlated with gene expression level (Stergachis et al. 2013). Taken together, our



**Figure 4.** Nucleosome occupancy profiles of 6-mer sites located in different genic and intergenic regions and their relationships to transcript abundance. (A) Heatmaps showing *A. thaliana* nucleosome occupancy around 6-mer motif sites located within distal promoter, proximal promoter, 5' untranslated region (5' UTR), CDS (coding sequences), intron, 3' UTR, and TTS 3' (500 bp downstream from TTS) regions. The 4096 6-mers (*y*-axis) were sorted according to normalized NOCs at the regions centered on the motif sites with those with lower scores (lower nucleosome occupancy) at the *bottom*. For distal promoter, only divergent genes where the TSS is 1.5 kb away from the TSS of the upstream gene were considered. For TTS 3' regions, only convergent genes where the TSSs are 500 bp away from the TSS of the downstream gene were included. (Arrowhead) 6-mer motif position. (B) Rice 6-mer motif occupancy profiles in different regions as in A. (C) Relationships between median normalized NOC scores of each 6-mer and median transcript abundance of genes with the 6-mer in question located in *A. thaliana* proximal, distal, CDS, or TTS 3' regions. (D) Same as C except rice data are used. ( $\rho$ ) Spearman's rank correlation coefficient.

findings indicate that nucleosome-depletion associated with specific sequence motifs is correlated with the level of gene expression. In addition, regulatory motifs in CDS are likely a common feature among eukaryotes.

#### Nucleosome-phasing and position bias of known TFBSs

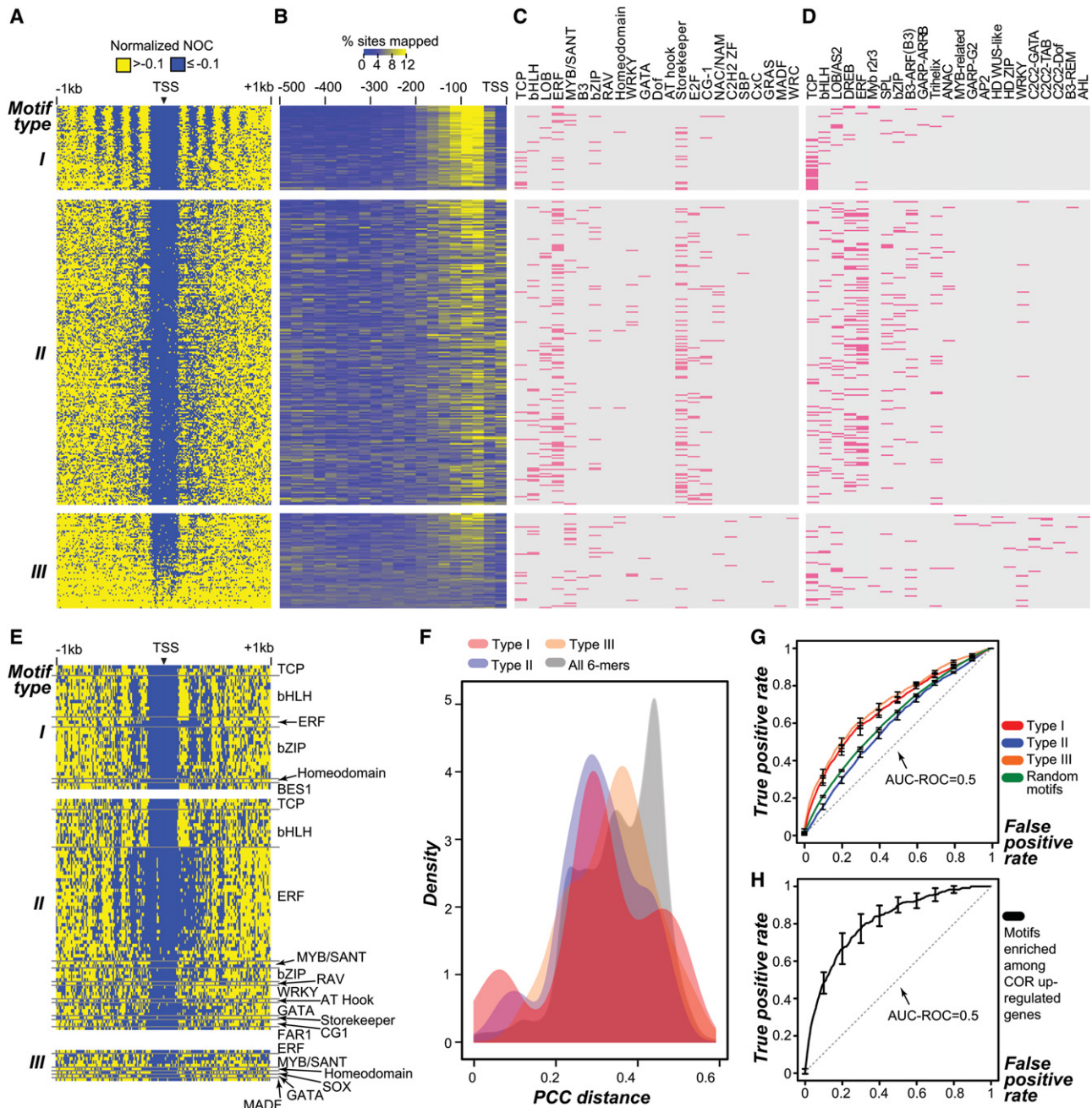
The correlation between motif-specific nucleosome depletion and expression level (Fig. 4) suggests that specific sequence motifs and their nucleosome occupancy jointly control gene expression. In

yeast, insertion of nucleosome-disfavoring sequences around TFBS reduces the nucleosome occupancy, increases DNA accessibility, and thus induces transcriptional activity (Raveh-Sadka et al. 2012). Consistent with this, 125 6-mer motifs that tend to be found in the proximal promoters of the top 10% most highly expressed genes (Fisher's exact test, adjusted  $P < 0.05$ ) (Methods) also have significantly lower median normalized NOC scores compared to other motifs (Kolmogorov-Smirnov tests,  $P < 2.2 \times 10^{-16}$ ), i.e., the NOC score distribution of the 125 6-mer motifs enriched in highly expressed genes and the distribution of all 6-mer motifs are significantly different. Among these motifs, 47% (59 of 125) that are overrepresented in highly expressed genes are significantly nucleosome-depleted (Type I motifs) (Fig. 3A, red-filled circles; Supplemental Table S2; Supplemental Fig. S5), raising the question whether these 59 Type I motifs are recognized by TFs. Genomic regions around Type I motifs display a strong phasing pattern (Fig. 5A, top), consistent with the phasing pattern around TFBSs in yeast, human, and rice (Mavrich et al. 2008; Valouev et al. 2011; Wu et al. 2014). In addition, Type I motifs display a strong positional bias in the proximal promoter regions (Fig. 5B, top), similar to known TFBSs in plants (Kaufmann et al. 2009; Shamimuzzaman and Vodkin 2013; Heyndrickx et al. 2014; Miyamoto et al. 2014). Note that this phasing pattern is not simply due to NDR because the normalized NOC score has been normalized to the background occupancy (Methods).

These findings reinforce the notion that motif sites satisfying both criteria, enriched in highly expressed genes and nucleosome-depleted, have a strong phasing pattern/position bias reminiscent of TFBSs. Consistent with this interpretation, the PCC distances (see Methods) between Type I motifs and their top matching known TF binding motifs (TFBMs) (Franco-Zorrilla et al. 2014) are significantly smaller than the distances

between all 6-mer motifs and their top matching TFBMs (Mann-Whitney  $U$  test,  $P$ -value  $< 1 \times 10^{-2}$ ) (Fig. 5F). Thus, Type I motifs are more similar to known TFBMs than most other 6-mers. In addition, most Type I motifs are similar to in vitro binding sites of TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR1 (TCP) family TFs (Fig. 5C,D, top; Supplemental Tables S8, S9). This is consistent with the fact that we used leaf DNA samples and earlier studies indicating TCP members' roles in leaf development (Iwakawa et al. 2007; Martín-Trillo and Cubas 2010). Aside from Type I motifs, 213 motifs have significantly lower





**Figure 5.** Properties of Type I, II, III motifs. (A) Nucleosome phasing pattern in 1-kb regions flanking 6-mer motif sites. Type I, II, and III motifs are as defined in Figure 3A. (Arrowhead) 6-mer motif position; (blue) lower occupancy; (yellow) higher occupancy. (B) Positional bias of motif sequences located within proximal promoter regions. Color represents the proportion (%) of 6-mer sites in a 25-bp bin compared to the entire region. (C) Families with TFs that likely bind to Type I, II, and III motifs. (Pink) 6-mer with significantly higher in vitro binding affinity ( $P \leq 0.01$ ) to a TF in a family (Weirauch et al. 2014) compared to random 6-mers (Methods). (D) TF families that likely bind to Type I, II, and III 6-mer motifs. (Pink) the best matching binding motif from a TF family (Franco-Zorrilla et al. 2014) to a 6-mer with PCC distance  $\leq 0.38$  ( $P < 0.05$ ) (Methods). (E) Nucleosome phasing pattern in 1-kb regions flanking TFBM sites (Weirauch et al. 2014). (F) PCC distance distributions of Type I, II, and III motifs and all 6-mers compared to their top matching TFBM (Franco-Zorrilla et al. 2014). (G) Performance of gene expression predictive models based on Type I, II, and III as well as randomly chosen 6-mer motifs. (Random guess) gray dotted line. (H) Performance of gene expression predictive models based on motifs enriched among COR-induced genes (black line). (Random guess) gray dotted line; (error bar) the standard error at 0.1 false positive rate intervals; (AUC-ROC) Area Under Curve-Receiver Operating Characteristic.

nucleosome occupancy relative to the background but are not significantly overrepresented among highly expressed genes (Type II) (Fig. 3A, blue circles; Fig. 5; Supplemental Table S2). Similar to Type I motifs, Type II motifs have significantly lower nucleosome

occupancy than background, a tendency to be located in proximal promoter regions (Fig. 5A,B, middle), and similar to Type I motifs, significantly higher sequence similarity to known TFBM (Mann-Whitney  $U$  test,  $P$ -value  $< 9.7 \times 10^{-12}$ ) (Fig. 5F). However, genomic

regions surrounding Type II motifs have only relatively weak nucleosome phasing, and positional preference is more diffused compared to Type I motifs (Fig. 5A,B, middle). Another 66 motifs enriched among highly expressed genes did not pass the stringent threshold to be classified as motifs with significant nucleosome depletion (Type III) (Fig. 3A, orange circles; Supplemental Table S2). Type III motifs have the highest nucleosome occupancy and the most diffused distribution (Fig. 5A,B, bottom). In contrast to Type I/II motifs, Type III motifs are not significantly more similar to TFBMs compared to all 6-mer motifs (Mann-Whitney  $U$  test,  $P$ -value = 0.07) (Fig. 5F), despite the fact that Type III motif sites are enriched among highly expressed genes (Fig. 3A; Supplemental Table S2). Note that the matching TFBM can be the same for multiple 6-mers (Fig. 5C,D); thus, it is not surprising that the 6-mers can be organized into multiple clusters, where each cluster contains 6-mers that may be bound by the same TFs (Supplemental Fig. S6; Supplemental Table S2).

Two TFBM data sets from *in vitro* binding studies (Franco-Zorrilla et al. 2014; Weirauch et al. 2014) were used in this analysis. The similar mapping patterns of Type I/II 6-mer motifs to multiple TF family members between these studies further suggest that these 6-mer motifs are potential TFBMs (Fig. 5C,D). To further demonstrate that nucleosome-depleted 6-mer motifs are likely TFBMs, we next examined *in vivo* binding data of 27 TFs (Heyndrickx et al. 2014). We found that, despite differences in growth and developmental conditions between the *in vivo* binding and our studies, the chromatin immunoprecipitation signal peaks of bHLHs and ERFs tend to overlap with Type I and II motif sites compared to randomly chosen sites (Fisher's exact test,  $P < 1 \times 10^{-4}$ ) (Supplemental Fig. S7; Methods). To assess the nucleosome occupancy pattern based on known TF binding motifs directly instead of 6-mers that may or may not be bound by TFs, we first classified TFBMs into three types based on the same criteria as that for classifying 6-mers (Fig. 5A). As expected (Fig. 5A,C,D), we found that the Type I/II motif sites of various TFBMs, particularly those relevant to TCP, bHLH, and ERF families, showed significant nucleosome depletion and strong phasing pattern (Fig. 5E, top and middle), supporting the observations in Figure 5A. These results are consistent with our interpretation that some Type I/II 6-mer motifs may be bound by TFs.

### Gene expression predictions based on 6-mer motifs and nucleosome occupancy

The three types of motifs have contrasting nucleosome depletion and phasing pattern, distribution within promoters, and correlations with highly expressed genes (Fig. 5A,B). To address the question of how well each type of motif may explain gene expression, we used the support vector machine (SVM) algorithm to establish models with motifs of each motif type (Fig. 5G) to predict whether a gene would be expressed at high or low level (Methods). We found that Type I motifs performed better (Area Under Curve-Receiver Operating Characteristic, AUC-ROC = 0.68) than either random guess (AUC-ROC = 0.5) or randomly picked 6-mers (AUC-ROC = 0.62) (Fig. 5G). Together with our findings on the abundance of Type I motifs in proximal promoter regions and their strong phasing pattern (Fig. 5A,B, top), the expression prediction with Type I motifs further indicate their relevance to gene expression regulation.

The model based on Type II motifs can only predict highly expressed genes with an AUC-ROC = 0.59 (Fig. 5G), in fact worse than the model based on randomly drawn 6-mer motifs (AUC-

ROC = 0.62) (Fig. 5G). This result implies that these Type II motifs are not strongly associated with gene expression, at least in the leaf tissue we analyzed. On the contrary, the model based on Type III motifs showed the best performance (AUC-ROC = 0.7) (Fig. 5G), even slightly better than the model based on Type I motifs. This is intriguing because Type I motifs are expected to be the most useful for predicting genes with high expression levels due to their location, nucleosome depletion, and strong phasing. Although Type III motifs are enriched in the proximal promoters of highly expressed genes, they are associated with relatively closed chromatin (Fig. 5A, bottom). One explanation is that some Type III motifs are bound by transcriptional repressors. For example, some Type III motifs resemble *in vitro* binding sites of C2C2-GATA and C2C2-YAB TFs ( $P$ -values < 0.05) (Fig. 5D, bottom) that may act as repressors (Richter et al. 2010; Franco-Zorrilla et al. 2014). Alternatively, some *trans*-acting factors may bind to their binding sites regardless of chromatin status as shown in studies of *A. thaliana* APETALA1 and SEPALLATA3 (Pajoro et al. 2014). Some TFs have also been shown to preferentially bind nucleosome-occupied regions (Teif et al. 2012; Ballare et al. 2013). Another consideration is that although some of the Type III motifs have higher nucleosome occupancy compared to Type I and II motifs, the occupancy in Type III motifs is still lower when all motifs are considered (Fig. 3A, orange dots versus all dots). Thus, it is possible that some Type III motifs are associated with altered nucleosomes, which have been shown to permit TF binding in yeast (Floer et al. 2010). Taken together, our findings show that lower nucleosome occupancy on specific motif sequences is an important predictor of gene expression level. In addition to motifs located within relatively open chromatin, those located in regions with higher nucleosome occupancy contribute to the gene expression prediction model.

### Influence of environmental condition on nucleosome occupancy

Our results show that the motif-associated nucleosome depletion is correlated with gene expression in *A. thaliana* leaves (Fig. 4C). In addition, many *cis*-regulatory elements are known to be responsible for dynamic changes of gene expression in various environments (Marcotte et al. 1989; Mundy et al. 1990; Baker et al. 1994; Zou et al. 2011; Hickman et al. 2013; Benn et al. 2014; Franco-Zorrilla et al. 2014; Sullivan et al. 2014). This raises the question of whether environmental stimuli may affect nucleosome occupancy on *cis*-regulatory elements and ultimately transcriptional response to environmental perturbation. To address this, we treated *A. thaliana* with coronatine (COR), which is an analog of the plant hormone jasmonoyl-L-isoleucine (JA-Ile) involved in regulation of gene expression in response to wounding (Feys et al. 1994; Uppalapati et al. 2005; Katsir et al. 2008; Koo and Howe 2009). We identified 366 genes as being significantly induced ( $\geq$  twofold increase in transcript level, adjusted  $P < 0.05$ ) within 1 h of COR treatment (Supplemental Fig. S8B). To assess whether COR-induced gene expression is correlated with changes in nucleosome occupancy, we first compared nucleosome occupancy changes between COR-induced and all annotated genes. Surprisingly, NOC scores before and after COR treatment were similar between COR-induced and annotated genes in all four genomic regions examined (250 bp upstream of TSS, TSS to 250 bp downstream, TSS to TTS, and 300 bp upstream of TTS) (Supplemental Fig. S8C). Thus, changes in nucleosome occupancy as measured in our approach were not apparently correlated with differential gene expression. Nonetheless, because nucleosome

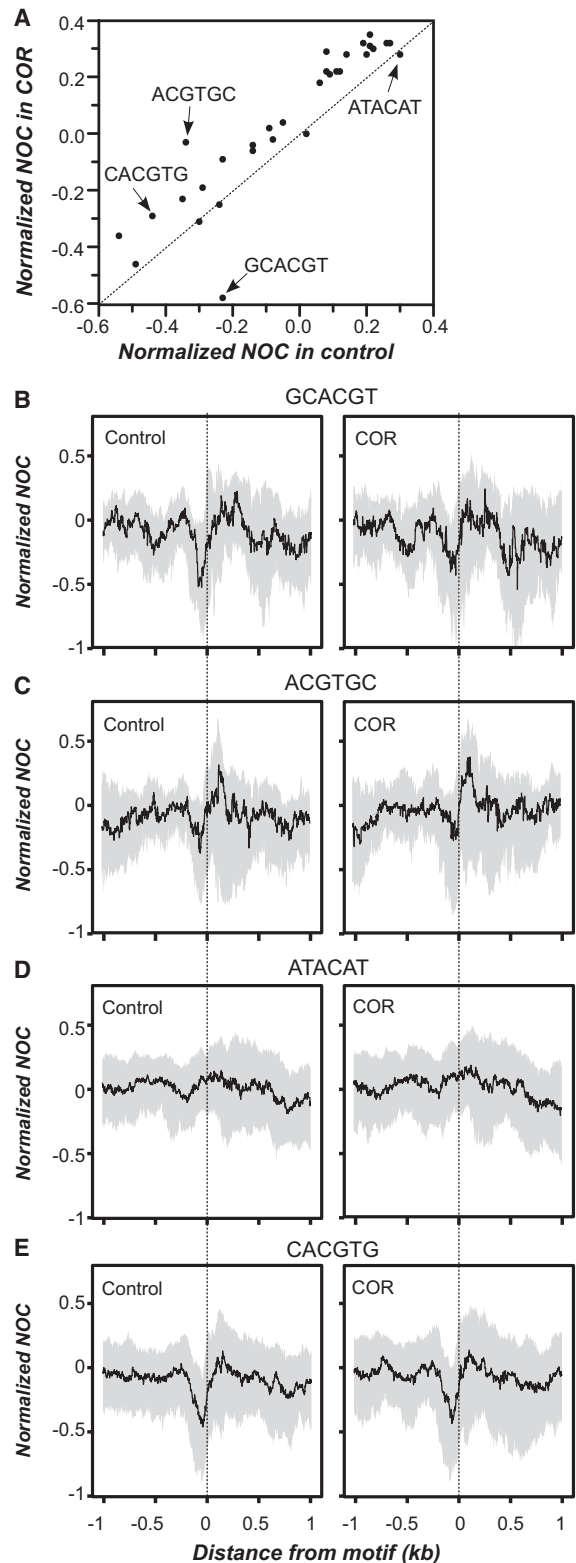


occupancy on 6-mer motifs is negatively correlated with gene expression levels (Figs. 4, 5), it is possible that occupancy differences are mostly in regions with *cis*-elements controlling the COR response instead of the entire genes and their surrounding sequences. To assess this possibility, we identified 31 6-mer motifs significantly enriched in proximal promoter regions of COR-induced genes compared to all genes (one-sided Fisher's exact test,  $P < 0.01$ ) (Supplemental Table S3), including the G-box known to regulate JA-responsive gene expression (Boter et al. 2004; Godoy et al. 2011). To demonstrate the relevance of these motifs to COR-induced gene expression, we applied the SVM algorithm to generate a COR-inducible gene prediction model based on these 31 enriched motifs. The 31 motif model performed significantly better (AUC-ROC=0.82) (Fig. 5H) than random guess (AUC-ROC=0.5), indicating these motifs are likely relevant to regulating the expression level of COR-responsive genes.

Next, we determined how COR treatment affected nucleosome occupancy in regions harboring these 31 motif sequences. The median normalized NOCs of some 6-mer motifs enriched in COR-induced genes are only slightly different between control and COR-treated samples (Fig. 6A; Supplemental Table S3). For example, the GCACGT and ACGTGC motifs are more nucleosome-depleted and nucleosome-enriched in COR versus control treatments, respectively (Fig. 6B,C). However, such difference is not significant statistically. In fact, none of the 31 motifs exhibited significantly different nucleosome occupancy level before and after COR treatment (Supplemental Table S3). Thus, the nucleosome occupancy levels on these motifs are indistinguishable before and after COR treatment. Interestingly, some of these motif sequences showed either nucleosome-depleted or enriched patterns (Supplemental Table S3). For example, the ATACAT motifs enriched in COR-induced genes were nucleosome-enriched in both control and COR samples (Fig. 6D; Supplemental Table S3). The ATACAT motif is similar to the AT-hook containing TF 20 (AHL20) (PCC distance = 0.24) (Methods), which has been implicated in the regulation of plant defense response (Vom Endt et al. 2007; Lu et al. 2010). In contrast to ATACAT motif, the G-box (CACGTG) showed a nucleosome-depleted pattern both in control and COR samples (normalized NOC = -0.49 in control and -0.46 in COR) (Fig. 6E; Supplemental Table S3), suggesting that G-box may be associated with TFs even in the absence of COR treatment. Consistent with this hypothesis, there is evidence to suggest that MYC-MYB TF complexes can bind to G-box motifs in the promoters of glucosinolate biosynthesis genes both in the absence and presence of the JA-Ile signal (Schweizer et al. 2013). In this case, additional regulators of JA-mediated defense responses may initiate rapid changes in gene expression without the need to displace nucleosomes (Howe 2010). Taken together, these results suggest that the DNA accessibility of *cis*-regulatory motifs in plant JA-response genes may be poised to facilitate a rapid transcriptional response to environmental stimuli.

## Conclusion

Through genome-wide analysis of nucleosome occupancy, we revealed that the NDR in promoters would explain 22% variation of gene expression. Beyond that, we further characterized the chromatin state around 6-mer motif sequences and demonstrated their relevance to gene expression in *A. thaliana*. We found significant negative correlations between G/C content (genomic regions and 6-mer motifs) and nucleosome occupancy in *A. thaliana* and rice that are distinct from budding yeast. In addition, after control



**Figure 6.** Nucleosome occupancy of COR-induced genes. (A) Comparison of normalized NOC scores for 6-mers enriched among coronatine (COR)-induced genes between control and COR-treated conditions. (B–E) The NOC (per base) profiles for GCACGT (B), ACGTGC (C), ATACAT (D), and CACGTG (E) motifs located in the proximal promoters of COR-induced genes under control and COR-treated conditions. The x-axis represents the distance to 6-mer sites.

for G/C content, we found that 6-mer motifs differ significantly in their nucleosome occupancy. Intriguingly, sites of a 6-mer motif that have nucleosome depletion in the proximal promoter region tend to have overall lower occupancy in other intergenic and genic regions. Furthermore, this sequence-specific nucleosome-depletion pattern is tightly associated with DHS. Because these patterns are conserved between *A. thaliana* and rice but not budding yeast, we suggest that these motif sites, including those located in promoter or coding sequences, may be part of a conserved mechanism for controlling plant gene expression. Consistent with this notion, regions containing a subset of the nucleosome-depleted motifs exhibit strong nucleosome-phasing patterns that are indicative of TF binding such as TCP, ERF, and bHLH families. This is further supported by our ability to distinguish highly and lowly expressed genes in machine learning predictions using motif information and nucleosome occupancy. A recent study provided an extensive look at the genome-wide, tissue-specific, and condition-specific DHSs and identified known/novel motifs in open chromatin regions in *A. thaliana* (Sullivan et al. 2014). Our findings complement existing nucleosome profiling and DHS studies by providing an explicit link between chromatin architecture, sequence motifs, and gene expression.

Although nucleosome-depleted regions around some 6-mer motif sequences can partly explain gene expression levels, this is not always the case. In addition, contrary to our expectation that induced gene expression should have significant changes in nucleosome occupancy, motifs enriched among COR-induced genes showed no clear difference in nucleosome occupancy before and after treatment. One explanation for these findings is that these motifs are accessible to TFs both before and after perception of environmental cues that trigger JA production. This would be similar to *Drosophila* heat shock TF (Guertin and Lis 2010) and the mammalian glucocorticoid receptor (John et al. 2011) that bind to sites that are accessible prior to heat shock and glucocorticoid treatments, respectively. Potentially *trans* factors in addition to those already bound may be required to kick-start transcriptional initiation. This is consistent with the findings that the number of TFBSs far exceeds the number of regulated genes (MacQuarrie et al. 2011) and plant TF binding to target genes do not necessarily lead to transcriptional initiation (Para et al. 2014). We also found that some motif sites have higher nucleosome occupancy compared to those with clear phasing patterns but are enriched among and predictive of highly expressed gene. Thus, despite the fact that we are able to classify highly and lowly expressed genes based on nucleosome occupancy alone, there are additional players in how chromatin states contribute to plant gene expression regulation that are yet to be identified.

In addition to regulatory components relevant to plant transcriptional regulation, there are a few limitations in our approach. First, because rosette leaf tissue used in this study contains multiple cell types, the nucleosome occupancy we revealed may mask cell-type-specific nucleosome occupancy patterns. Second, instead of using one concentration of MNase, digestion of gDNA with a wide-range of MNase concentrations will provide a better kinetic understanding of chromatin accessibility, particularly in identifying weakly bound nucleosomes and subtle changes in nucleosome occupancy in different environments (Bryant et al. 2008). Despite these shortcomings, our findings advance the understanding of the impact of chromatin structure on the regulation of plant gene expression. We demonstrate that sequence-specific nucleosome depletion in promoter and CDS regions is an informative and evolutionarily conserved feature for gene expression pre-

diction. We also provide quantitative assessments of nucleosome occupancy among specific motif sequences on plant gene expression. The predictive models described in this study provide a baseline on which to incorporate additional factors controlling transcriptional regulation.

## Methods

### Plant materials and isolation of DNA and RNA

*Arabidopsis thaliana* Col-0 seeds were imbibed in water at 4°C in darkness for 4 d for stratification, grown on soil (Arabidopsis mix) under long-day conditions (16 h of light/8 h of dark) at a fluence rate of 100  $\mu\text{mol}/\text{m}^2/\text{sec}^1$ , and watered with half-strength Hoagland nutrient solution once per week (otherwise with deionized water). After ~4 wk, the aerial tissues before bolting were harvested and immediately immersed in liquid nitrogen. To assess the effect of environmental stimuli on nucleosome occupancy, plants were treated with coronatine (COR) (Sigma) or water (control), as described previously (Attaran et al. 2014). The harvested samples were ground with liquid nitrogen and divided into three aliquots for nucleosome genomic DNA (gDNA), naked gDNA, and RNA isolation. For isolating nucleosome gDNA, the nuclei were purified from one aliquot with nuclei extraction buffer (10 mM Hepes [pH 7.6], 5 mM potassium chloride, 5 mM magnesium chloride, 1 M sucrose, 14 mM  $\beta$ -mercaptoethanol, 0.6% Triton X-100, 0.4 mM phenylmethanesulfonyl fluoride, 1 $\times$  ethylenediamine-tetraacetic acid-free protease inhibitor [Roche]), digested with 0.4 units/ $\mu\text{L}$  micrococcal nuclease (MNase) (NEB) for 9 min, and purified with phenol/chloroform. To generate the naked gDNA sample, the gDNA was isolated without the nucleus isolation step as described previously (Dellaporta et al. 1983), purified with phenol/chloroform to strip off proteins, and digested with 0.25 units/ $\mu\text{L}$  MNase (NEB) for 1.75 or 3 min. Instead of cutting the 150 bp DNA fragments from the agarose gels (Chodavarapu et al. 2010), all the purified DNA fragments were used to construct libraries for 50-bp paired-end (PE) sequencing, following an approach that evaluates not only nucleosome occupancy but also regulator binding sites (Henikoff et al. 2011). For assessing the relationships between nucleosome occupancy and expression, total RNA was purified with the E.Z.N.A. Plant RNA kit (Omega Bio-Tek) and used to construct a single-end sequencing library following the Illumina protocol.

Four RNA samples (two control and two COR-treated samples) were indexed separately and then pooled into one lane of a flow-cell. Four nucleosome gDNA samples (two control and two COR-treated samples) and one naked gDNA sample were indexed separately, pooled together, and all five indexed samples were loaded into each of two lanes on one flow cell. Sequencing was performed with the Illumina HiSeq 2500 platform. Approximately 150 million 50-bp reads were obtained per lane.

### Sequencing data processing

To map the MNase-seq and RNA-seq reads, *A. thaliana* reference genome sequence and gene models were retrieved from The *A. thaliana* Information Resource (TAIR10; <http://www.arabidopsis.org>). To focus on the analysis of mononucleosome protected sequences, MNase-seq reads from nucleosome-protected or naked gDNA were further analyzed if the fragment size (distance between PE reads) was between 140 and 170 bp. These qualified PE reads were trimmed with Trimmomatic (default setting except leading = 20, trailing = 20, and minlen = 20) (Bolger et al. 2014) and mapped to the genome allowing two mismatches with Bowtie (version 1.0.0) (Langmead et al. 2009). For each nucleotide in

the genome, the degree of nucleosome occupancy was represented by the Nucleosome OCcupancy (NOC) score that is the number of reads mapped to the position in question divided by the total mapped reads per million for each sample. For comparison, a second approach was used by calculating coverage of each genomic position considered only the midpoint position of each read (Fig. 5A,E; Supplemental Fig. S1D). Because the genome-wide nucleosome occupancy of control and COR sample replicates are highly correlated (Pearson's correlation coefficient  $>0.96$ ) (Supplemental Table S5), the reads from replicates were combined together for the downstream analyses. Note the biological variation of naked gDNA sample is unaccounted for because only one replicate was generated.

Transcript levels of annotated genes were determined by mapping the RNA-seq reads to the genome with TopHat (version 1.4.1) (Trapnell et al. 2009) and were shown as fragments per kilobase per million fragments mapped (FPKM) calculated with Cufflinks (version 2.2.0) (Trapnell et al. 2010). Only protein-coding genes with FPKM "low confidence interval"  $>0$  (Trapnell et al. 2010) were included in downstream analyses. To identify COR up-regulated genes, the transcript levels of control and COR-treated samples were compared with EdgeR (Robinson et al. 2010). Any gene with false discovery rate adjusted  $P < 0.05$  (Benjamini and Hochberg 1995) and  $\geq$ twofold increase in RNA level under the COR-treated condition was considered up-regulated. The numbers of raw, quality filtered, and mapped reads and the sequencing coverage were reported in Supplemental Table S4.

The *A. thaliana* DNase I sequencing data set used is from a published study examining leaf tissues of 2-wk-old plants (Zhang et al. 2012). The mapping method of DNase I-seq data was the same as for MNase-seq. To determine DNase I hypersensitivity (DHS), the positions of the DNase I cutting sites were inferred from the 5' coordinates of the mapped reads according to published studies (Boyle et al. 2008; Sullivan et al. 2014).

The same methods used for *A. thaliana* were also used to analyze the rice seedling and yeast MNase-seq and RNA-seq data sets. The genome sequences and gene models were based on the genome version release 7 for rice (<http://rice.plantbiology.msu.edu>) and R64-1-1 for budding yeast (<http://www.yeastgenome.org>). Because yeast gene annotations contain little or no information about intron and UTR coordinates, respectively, and the yeast genes are so densely distributed that few genes fit the distance/transcription-direction criteria for distal and TTS 3' analyses (Supplemental Table S1), these regions were not included in yeast nucleosome occupancy analyses.

### Calculation of the nucleosome positioning, DHS, and in vivo TF binding around motif sequences

Six-mer motif sequences occurring in genic as well as the flanking regions 1 kb upstream of and 500 bp downstream from protein coding genes were identified using a customized python script. The motif sites of the TFBMs on proximal promoters of genes in Figure 5E were identified with motility (<http://cartwheel.caltech.edu/motility/intro.html>) with  $P$ -value  $< 1 \times 10^{-4}$ . Only the protein-coding genes with FPKM "low confidence interval"  $>0$  (Trapnell et al. 2010) in control or COR samples were considered to be expressed and included in the further analyses. For each genomic position containing a 6-mer, a normalized "regional Nucleosome OCcupancy (NOC)" score was calculated by taking the  $\log_2$  ratio between the average nucleosome occupancy of a 200-bp genomic region centered on the motif site to those on the same coordinates of a randomly chosen gene. Because lower nucleosome occupancy was seen in the direction of transcription (Supplemental Fig. S2A), this normalization was to control for

the background occupancy, particularly considering the presence of NDR in the proximal promoter region. For each 6-mer sequence, the median of normalized NOC scores described above was used to represent the nucleosome occupancy surrounding the 6-mer in question. To identify *A. thaliana* 6-mer motifs that are significantly enriched in nucleosome-depleted or nucleosome-enriched regions (Fig. 3A), the Mann-Whitney  $U$  test was performed to evaluate the significance of the differences between the nucleosome occupancy levels of a 200-bp genomic region centered on the motif site to those on the same coordinates of a randomly chosen gene.

The DNase I hypersensitivity (DHS) score for each 6-mer motif site in Figure 3C,D was calculated by averaging the number of cut positions per million DNase I mapped reads within 150 bp centered on the motif site. For each 6-mer motif, the median DHS score among the motif sites was reported. In Figure 3E and Supplemental Figure S3A, DHS profiles represent the average DHS signal at each DNA base pair position among motif sites within a 2-kb window centered on the motif sequence. To determine whether 6-mer motifs is likely the in vivo binding motif of a TF (Supplemental Fig. S7), the degree of overlap between the 6-mer motif occurrences and chromatin immunoprecipitation (ChIP)-seq peaks (Heyndrickx et al. 2014) was determined. The statistical significance of overlap was evaluated using the Fisher's exact test. A ChIP-seq peak and a 6-mer motif site is considered overlapping if the midpoint of the region covered by a ChIP-seq peak is located within the 6-mer motif site.

### Six-mer motif enrichment among highly and lowly expressed genes and distances between 6-mers and known TFBMs

Genes were regarded as highly expressed if their expression level  $\geq$ 90th percentile FPKM values over all genes, whereas lowly expressed genes were those at the bottom 10th percentile. Enrichment of a 6-mer was determined by whether their occurrence in promoters of highly expressed genes is significantly higher than that in promoters of lowly expressed genes with the Fisher's exact test. Here, promoter was defined as the 500-bp region upstream of TSS. The test  $P$ -value was adjusted for multiple testing (Benjamini-Hochberg method) (Benjamini and Hochberg 1995). Only the 6-mers with adjusted  $P$ -value  $< 0.05$  were further analyzed (Fig. 5).

The 338 6-mers with significantly lower than expected degrees of nucleosome occupancy and/or enriched among highly expressed genes were compared to two sets of TFBMs: 279 TFBMs of 256 plant TFs from 30 families (Weirauch et al. 2014) and 108 TFBMs of 63 plant TFs from 26 families (Fig. 5C,D; Franco-Zorrilla et al. 2014). We analyzed the two data sets separately since the data available were different. In the smaller data set (Franco-Zorrilla et al. 2014), the distance measure between each 6-mer and a TFBM was calculated as 1-PCC (Pearson's correlation coefficient) where the PCC value was generated with TAMO (Gordon et al. 2005). To define that a 6-mer and a known TF binding motif were significantly more similar than expected between TF families (thus the 6-mer in question is likely bound by TF[s] from a family the 6-mer is similar to), the pairwise distances of known TF binding motifs across TF families were calculated, and the fifth percentile of distance, 0.38 (with a  $P$ -value = 0.05), was set as a threshold.

In the larger data set (Weirauch et al. 2014), the in vitro binding affinities of all the possible 8-mers of a TF in the Protein Binding Microarray (PBM) were available. The binding affinity of a 6-mer motif for the TF in question was estimated by taking the median value of the  $E$ -scores (Godoy et al. 2011) of all PBM



8-mers containing the 6-mer motif. The statistical significance of the estimated binding affinity was determined by bootstrapping with one million replicates. A 6-mer motif with an adjusted bootstrapping  $P$ -value  $< 0.01$  was considered to have significantly higher in vitro binding affinity to a TF. In Figure 5C, instead of indicating the specific TF, the family the TF in question belongs to was shown.

### Linear model for explaining expression levels based on nucleosome occupancy

To determine how nucleosome occupancy may explain expression level, a linear model was generated (with the `lm` function in R) (R Core Team 2014) in which the expression level (FPKM) of a gene was regressed based on the average NOC scores over positions in sequence bins from 1-kb regions flanking both the TSSs and TTSs. Genes with gene bodies smaller than 1 kb were excluded from the analysis because we examined 1-kb regions flanking TSSs and TTS, otherwise presumably genic bins would contain intergenic regions. To determine how the nucleosome occupancy in different regions relative to TSSs and TTSs may explain expression level differences among genes, the regions specified above were divided into equally sized bins of 50, 100, and 250 bp. We choose to use a 100-bp bin because it has a larger  $R^2$  (coefficient of determination, fraction of the gene expression explained) of the linear model compared to the 250-bp bin model and smaller variance inflation factors of the regression coefficient (the larger the factor is, the more serious a problem multicollinearity becomes) compared to the 50-bp bin model (Supplemental Table S6). The linear regression analysis described above was also performed on naked genomic DNA data. The results show that there might be a slight bias near TSS, but no strong effect on expression was identified (Fig. 1C).

### Gene expression prediction

We predict expression in two contexts with the support vector machine (SVM) method that allows modeling of gene expression considering the joint contributions of a set of motifs. In the first context, we predicted whether a gene would be highly or lowly expressed based on multiple sets of 6-mers. Here, the 6-mers significantly enriched among highly expressed genes and significantly nucleosome-depleted (Fig. 5) were used as attributes, whereas the binary expression status (lowly or highly expressed) was the class we wanted to predict. For comparison, 10 random sets of 272 6-mers from all possible 6-mers, excluding Type I, II and III motifs, were chosen for constructing background prediction models (Fig. 5G). For each gene, the presence or absence of enriched 6-mers within the 500-bp region upstream of TSS was used to predict the class of the gene.

In the second context, we predicted whether a gene would be induced by COR treatment or not. Here, the 6-mers significantly enriched among COR-induced genes were used as attributes for prediction. In both contexts, expression prediction was performed using the LIBSVM implementation of the SVM method through the Weka wrapper (Hall et al. 2009; Chang and Lin 2011). A grid search was used to identify the best parameter combinations (Supplemental Table S7). Each parameter combination was tested with 10-fold cross-validation. The model with the parameter combination that maximized the average AUC-ROC from the 10-fold runs was regarded as the best model. ROC curves were plotted using the `ROCR` package in R (Sing et al. 2005). The ROC curve for random 6-mer prediction represents the average of the 10-folds for all 10 random sets.

### Data access

The MNase- and RNA-seq data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE64397.

### Acknowledgments

We thank David Arnosti and members of the Shiu laboratory for discussion and Melissa Lehti-Shiu for helpful comments on the manuscript. This work was in part supported by the National Science Foundation grants (MCB-1119778 and IOS-1126998) and Michigan State University to S.-H.S.; by the Natural Sciences and Engineering Research Council of Canada to I.T.M.; the Michigan AgBioResearch Project (M1CL02278) to G.A.H.; and the Taiwan Ministry of Science and Technology (104-2917-I-564-070) to Z.T.-Y.T.

*Author contributions:* M.-J.L., M.F., G.A.H., and S.-H.S. designed the research. M.-J.L. and I.T.M. performed the experiments. M.-J.L., A.E.S., and Z.T.-Y.T. analyzed the data. M.-J.L., A.E.S., Z.T.-Y.T., I.T.M., M.F., G.A.H., and S.-H.S. wrote the paper.

### References

- Attaran E, Major IT, Cruz JA, Rosa BA, Koo AJ, Chen J, Kramer DM, He SY, Howe GA. 2014. Temporal dynamics of growth and photosynthesis suppression in response to jasmonate signaling. *Plant Physiol* **165**: 1302–1314.
- Baker SS, Wilhelm KS, Thomashow MF. 1994. The 5'-region of *Arabidopsis thaliana cor15a* has cis-acting elements that confer cold-, drought- and ABA-regulated gene expression. *Plant Mol Biol* **24**: 701–713.
- Ballare C, Castellano G, Gaviglia L, Althammer S, González-Vallinas J, Eyrales E, Le Dily F, Zaurin R, Soronellas D, Vicent GP, et al. 2013. Nucleosome-driven transcription factor binding and gene regulation. *Mol Cell* **49**: 67–79.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* **57**: 289–300.
- Benn G, Wang CQ, Hicks DR, Stein J, Guthrie C, Dehesh K. 2014. A key general stress response motif is regulated non-uniformly by CAMTA transcription factors. *Plant J* **80**: 82–92.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Boter M, Ruiz-Rivero O, Abdeen A, Prat S. 2004. Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and *Arabidopsis*. *Genes Dev* **18**: 1577–1591.
- Boyle AP, Guinney J, Crawford GE, Furey TS. 2008. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics* **24**: 2537–2538.
- Bryant GO, Prabhu V, Floer M, Wang X, Spagna D, Schreiber D, Ptashne M. 2008. Activator control of nucleosome occupancy in activation and repression of transcription. *PLoS Biol* **6**: 2928–2939.
- Chang CC, Lin CJ. 2011. LIBSVM: a library for support vector machines. *ACM Trans Intel Syst Technol* **2**: 27.
- Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, et al. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature* **466**: 388–392.
- Chung HR, Dunkel I, Heise F, Linke C, Krobitch S, Ehrenhofer-Murray AE, Sperling SR, Vingron M. 2010. The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS One* **5**: e15754.
- Dellaporta SL, Wood J, Hicks JB. 1983. A plant DNA miniprep: version II. *Plant Mol Biol Rep* **1**: 19–21.
- Feys B, Benedetti CE, Penfold CN, Turner JG. 1994. *Arabidopsis* mutants selected for resistance to the phytotoxin coronatine are male sterile, insensitive to methyl jasmonate, and resistant to a bacterial pathogen. *Plant Cell* **6**: 751–759.
- Fincher JA, Vera DL, Hughes DD, McGinnis KM, Dennis JH, Bass HW. 2013. Genome-wide prediction of nucleosome occupancy in maize reveals plant chromatin structural features at genes and other elements at multiple scales. *Plant Physiol* **162**: 1127–1141.
- Floer M, Wang X, Prabhu V, Berrozpe G, Narayan S, Spagna D, Alvarez D, Kendall J, Krasnitz A, Stepansky A, et al. 2010. A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding. *Cell* **141**: 407–418.

- Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R. 2014. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci* **111**: 2367–2372.
- Godoy M, Franco-Zorrilla JM, Pérez-Pérez J, Oliveros JC, Lorenzo O, Solano R. 2011. Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors. *Plant J* **66**: 700–711.
- Gordon DB, Neklyudova L, McCallum S, Fraenkel E. 2005. TAMO: a flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs. *Bioinformatics* **21**: 3164–3165.
- Gossett AJ, Lieb JD. 2012. *In vivo* effects of histone H3 depletion on nucleosome occupancy and position in *Saccharomyces cerevisiae*. *PLoS Genet* **8**: e1002771.
- Gross DS, Garrard WT. 1988. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* **57**: 159–197.
- Guertin MJ, Lis JT. 2010. Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet* **6**: e1001114.
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. 2009. The WEKA data mining software: an update. *SIGKDD Explor Newslett* **11**: 10–18.
- Henikoff JG, Belsky JA, Kravinsky K, MacAlpine DM, Henikoff S. 2011. Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci* **108**: 18318–18323.
- Heyndrickx KS, Van de Velde J, Wang C, Weigel D, Vandepoele K. 2014. A functional and evolutionary perspective on transcription factor binding in *Arabidopsis thaliana*. *Plant Cell* **26**: 3894–3910.
- Hickman R, Hill C, Penfold CA, Breeze E, Bowden L, Moore JD, Zhang P, Jackson A, Cooke E, Bewick-Copley F, et al. 2013. A local regulatory network around three NAC transcription factors in stress responses and senescence in *Arabidopsis* leaves. *Plant J* **75**: 26–39.
- Howe GA. 2010. Ubiquitin ligase-coupled receptors extend their reach to jasmonate. *Plant Physiol* **154**: 471–474.
- Hughes AL, Rando OJ. 2014. Mechanisms underlying nucleosome positioning *in vivo*. *Annu Rev Biophys* **43**: 41–63.
- Hughes AL, Jin Y, Rando OJ, Struhl K. 2012. A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Mol Cell* **48**: 5–15.
- Iwakawa H, Iwasaki M, Kojima S, Ueno Y, Soma T, Tanaka H, Semiarti E, Machida Y, Machida C. 2007. Expression of the *ASYMMETRIC LEAVES2* gene in the adaxial domain of *Arabidopsis* leaves represses cell proliferation in this domain and is critical for the development of properly expanded leaves. *Plant J* **51**: 173–184.
- Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F, bZIPRG. 2002. bZIP transcription factors in *Arabidopsis*. *Trends Plant Sci* **7**: 106–111.
- John S, Sabo PJ, Thurman RE, Sung MH, Biddie SC, Johnson TA, Hager GL, Stamatoyannopoulos JA. 2011. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* **43**: 264–268.
- Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**: 362–366.
- Katsir L, Chung HS, Koo AJ, Howe GA. 2008. Jasmonate signaling: a conserved mechanism of hormone sensing. *Curr Opin Plant Biol* **11**: 428–435.
- Kaufmann K, Muiño JM, Jauregui R, Airolidi CA, Smaczniak C, Krajewski P, Angenent GC. 2009. Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the *Arabidopsis* flower. *PLoS Biol* **7**: e1000090.
- Koo AJ, Howe GA. 2009. The wound hormone jasmonate. *Phytochemistry* **70**: 1571–1580.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* **39**: 1235–1244.
- Li G, Liu S, Wang J, He J, Huang H, Zhang Y, Xu L. 2014. ISWI proteins participate in the genome-wide nucleosome distribution in *Arabidopsis*. *Plant J* **78**: 706–714.
- Lu H, Zou Y, Feng N. 2010. Overexpression of *AHL20* negatively regulates defenses in *Arabidopsis*. *J Integr Plant Biol* **52**: 801–808.
- MacQuarrie KL, Fong AP, Morse RH, Tapscott SJ. 2011. Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet* **27**: 141–148.
- Marcotte WR Jr, Russell SH, Quatrano RS. 1989. Abscisic acid-responsive sequences from the Em gene of wheat. *Plant Cell* **1**: 969–976.
- Martín-Trillo M, Cubas P. 2010. TCP genes: a family snapshot ten years later. *Trends Plant Sci* **15**: 31–39.
- Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF. 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* **18**: 1073–1083.
- Miyamoto K, Matsumoto T, Okada A, Komiyama K, Chujo T, Yoshikawa H, Nojiri H, Yamane H, Okada K. 2014. Identification of target genes of the bZIP transcription factor OsTGAP1, whose overexpression causes elicitor-induced hyperaccumulation of diterpenoid phytoalexins in rice cells. *PLoS One* **9**: e105823.
- Mundy J, Yamaguchi-Shinozaki K, Chua NH. 1990. Nuclear proteins bind conserved elements in the abscisic acid-responsive promoter of a rice *rab* gene. *Proc Natl Acad Sci* **87**: 1406–1410.
- Pajoro A, Madrigal P, Muiño JM, Matus JT, Jin J, Mecchia MA, Debernardi JM, Palatnik JF, Balazadeh S, Arif M, et al. 2014. Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol* **15**: R41.
- Para A, Li Y, Marshall-Colón A, Varala K, Francoeur NJ, Moran TM, Edwards MB, Hackley C, Bargmann BO, Birnbaum KD, et al. 2014. Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in *Arabidopsis*. *Proc Natl Acad Sci* **111**: 10371–10376.
- R Core Team. 2014. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Raveh-Sadka T, Levo M, Shabi U, Shany B, Keren L, Lotan-Pompan M, Zeevi D, Sharon E, Weinberger A, Segal E. 2012. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat Genet* **44**: 743–750.
- Richmond TJ, Davey CA. 2003. The structure of DNA in the nucleosome core. *Nature* **423**: 145–150.
- Richter R, Behringer C, Müller IK, Schwecheimer C. 2010. The GATA-type transcription factors GNC and GNL/CGA1 repress gibberellin signaling downstream from DELLA proteins and PHYTOCHROME-INTERACTING FACTORS. *Genes Dev* **24**: 2093–2104.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Schweizer F, Fernández-Calvo P, Zander M, Diez-Diaz M, Fonseca S, Glauser G, Lewsey MG, Ecker JR, Solano R, Reymond P. 2013. *Arabidopsis* basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. *Plant Cell* **25**: 3117–3132.
- Shamimuzzaman M, Vodkin L. 2013. Genome-wide identification of binding sites for NAC and YABBY transcription factors and co-regulated genes during soybean seedling development by ChIP-Seq and RNA-Seq. *BMC Genomics* **14**: 477.
- Sing T, Sander O, Beerenwinkel N, Lengauer T. 2005. ROCr: visualizing classifier performance in R. *Bioinformatics* **21**: 3940–3941.
- Stergachis AB, Haugen E, Shafer A, Fu W, Vernot B, Reynolds A, Raubitschek A, Ziegler S, LeProust EM, Akey JM, et al. 2013. Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* **342**: 1367–1372.
- Struhl K, Segal E. 2013. Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**: 267–273.
- Sullivan AM, Arsovski AA, Lempe J, Bubbs KL, Weirauch MT, Sabo PJ, Sandstrom R, Thurman RE, Neph S, Reynolds AP, et al. 2014. Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. *Cell Rep* **8**: 2015–2030.
- Sung MH, Guertin MJ, Baek S, Hager GL. 2014. DNase footprint signatures are dictated by factor dynamics and DNA sequence. *Mol Cell* **56**: 275–285.
- Teif VB, Vainshtein Y, Caudron-Herger M, Mallm JP, Marth C, Höfer T, Rippe K. 2012. Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol* **19**: 1185–1192.
- Toledo-Ortiz G, Huq E, Quail PH. 2003. The *Arabidopsis* basic/helix-loop-helix transcription factor family. *Plant Cell* **15**: 1749–1770.
- Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.
- Uppalapati SR, Ayoubi P, Weng H, Palmer DA, Mitchell RE, Jones W, Bender CL. 2005. The phytotoxin coronatine and methyl jasmonate impact multiple phytohormone pathways in tomato. *Plant J* **42**: 201–217.
- Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. 2011. Determinants of nucleosome organization in primary human cells. *Nature* **474**: 516–520.
- Vera DL, Madzima TF, Labonne JD, Alam MP, Hoffman GG, Girimurugan SB, Zhang J, McGinnis KM, Dennis JH, Bass HW. 2014. Differential nucleosome sensitivity profiling of chromatin reveals biochemical footprints coupled to gene expression and functional DNA elements in maize. *Plant Cell* **26**: 3883–3893.

- Vierstra J, Wang H, John S, Sandstrom R, Stamatoyannopoulos JA. 2014. Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. *Nat Methods* **11**: 66–72.
- Vom Endt D, Soares e Silva M, Kijne JW, Pasquali G, Memelink J. 2007. Identification of a bipartite jasmonate-responsive promoter element in the *Catharanthus roseus* ORCA3 transcription factor gene that interacts specifically with AT-Hook DNA-binding proteins. *Plant Physiol* **144**: 1680–1689.
- Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y, et al. 2012. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* **22**: 1798–1812.
- Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**: 1431–1443.
- Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc Natl Acad Sci* **86**: 6201–6205.
- Wu Y, Zhang W, Jiang J. 2014. Genome-wide nucleosome positioning is orchestrated by genomic regions associated with DNase I hypersensitivity in rice. *PLoS Genet* **10**: e1004378.
- Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**: 626–630.
- Zhang Z, Wippo CJ, Wal M, Ward E, Korber P, Pugh BF. 2011. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* **332**: 977–980.
- Zhang W, Zhang T, Wu Y, Jiang J. 2012. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. *Plant Cell* **24**: 2719–2731.
- Zou C, Sun K, Mackaluso JD, Seddon AE, Jin R, Thomashow MF, Shiu SH. 2011. Cis-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*. *Proc Natl Acad Sci* **108**: 14992–14997.

Received December 19, 2014; accepted in revised form June 10, 2015.