DNA-binding specificities of plant transcription factors and their potential to define target genes

José M. Franco-Zorrilla[a,1], Irene López-Vidriero[a], José L. Carrasco[b], Marta Godoy[a], Pablo Vera[b], and Roberto Solano[c,1]

[a]Genomics Unit and [c]Department of Plant Molecular Genetics, Centro Nacional de Biotecnología (CNB)-Consejo Superior de Investigaciones Científicas (CSIC), Darwin 3, 28049 Madrid, Spain; and [b]Instituto de Biología Molecular y Celular de Plantas, Universidad Politécnica de Valencia-Consejo Superior de Investigaciones Científicas, 46022 Valencia, Spain

Transcription factors (TFs) regulate gene expression through binding to *cis*-regulatory specific sequences in the promoters of their target genes. In contrast to the genetic code, the transcriptional regulatory code is far from being deciphered and is determined by sequence specificity of TFs, combinatorial cooperation between TFs and chromatin competence. Here we addressed one of these determinants by characterizing the target sequence specificity of 63 plant TFs representing 25 families, using protein-binding microarrays. Remarkably, almost half of these TFs recognized secondary motifs, which in some cases were completely unrelated to the primary element. Analyses of coregulated genes and transcriptomic data from TFs mutants showed the functional significance of over 80% of all identified sequences and of at least one target sequence per TF. Moreover, combining the target sequence information with coexpression analysis we could predict the function of a TF as activator or repressor through a particular DNA sequence. Our data support the correlation between *cis*-regulatory elements and the sequence determined in vitro using the protein-binding microarray and provides a framework to explore regulatory networks in plants.

Arabidopsis | regulatory network | transcriptional regulation | *cis*-element | coregulation

Transcription factors (TFs) mediate cellular responses through recognizing specific *cis*-regulatory DNA sequences at the promoters of their targets genes. In plants, organ development is a continuous process that expands beyond the embryonic phase and, as sessile organisms, plants have to face with a wide range of environmental stresses. Signaling cascades governing developmental and stress switches converge at the gene expression level. Pioneering work (1) suggested that transcriptional regulation may play more important roles in plants than in animals, given the large number of TF-coding genes in plant genomes, ranging from 6% to 10%, depending on the database.

During the last few years, the advance in the determination of TF-binding sites, including both in vivo and in vitro techniques, is helping to decipher the transcriptional regulatory code (2–4). In vivo approaches involving immunoprecipitation of TF-bound chromatin followed by microarray or sequencing analysis (ChIP-chip and ChIP-seq, respectively) are contributing to the knowledge of the transcriptional networks associated with a TF. ChIP-based techniques revealed that TFs may bind to thousands of genomic fragments, suggesting that the TF is interacting indirectly with DNA or that the binding requires additional cooperative factors (5, 6). The situation may not be different in the case of plant genomes. To date, only a limited number of studies have deepened in the discovery of the targets of some TFs and found that, similar to TFs in animals, TFs in plants bind to hundreds or thousands of DNA fragments; in some cases, only a small proportion of targets respond transcriptionally to the TF, obscuring the identification of actual binding sites (7–9). In this context, the precise identification of the DNA-binding sequence of each TF may be instrumental to clarify the transcriptional regulatory code and to allow development of predictive models of transcriptional regulation.

The application of high-throughput in vitro techniques is making the identification of the binding-sequences of all of the TFs in a genome an affordable task. SELEX-seq and protein-binding microarrays (PBMs) have yielded information of binding motifs for hundreds of TFs in mammals but these studies still lack of a simple and systematic analysis of the biological relevance of the motifs (3, 4). In this study, we defined the DNA-binding motifs for 63 *Arabidopsis thaliana* TFs in vitro by means of PBM analysis, with a particular emphasis of plant-specific families, and observed that approximately half of them may recognize secondary motifs. By analyzing coregulated genes, we found significant biological relevance for at least one binding motif for all of the TFs analyzed and for more than 80% of the identified motifs. These results indicate that binding sequences obtained in vitro coupled with analysis of coregulated genes provide useful functional information for the identification of *cis*-regulatory elements and TF-target genes, which complements ChIP-seq data and offers a framework for an easy and systematic analysis of the regulatory networks in plants.

Results and Discussion

Characterization of DNA-Binding Specificity of 63 *A. thaliana* TFs. We cloned ~100 TFs as N-terminal Maltose Binding Protein (MBP)-tagged fusions for expression in *Escherichia coli* and analyzed their DNA-binding specificity by incubation of PBMs (10). We obtained specific DNA-binding sequences for 63 TFs (Fig. 1), representing 2.5–4.5% of the total complement of the TFs in *A. thaliana*. This dataset includes 25 families or subfamilies of TFs, 15 of them plant-specific (Fig. 1; *SI Appendix*, Tables S1 and

---

Significance

We described the high-throughput identification of DNA-binding specificities of 63 plant transcription factors (TFs) and their relevance as *cis*-regulatory elements in vivo. Almost half of the TFs recognized secondary motifs partially or completely differing from their corresponding primary ones. Analysis of coregulated genes, transcriptomic data, and chromatin hypersensitive regions revealed the biological relevance of more than 80% of the binding sites identified. Our combined analysis allows the prediction of the function of a particular TF as activator or repressor through a particular DNA sequence. The data support the correlation between *cis*-regulatory elements in vivo and the sequence determined in vitro. Moreover, it provides a framework to explore regulatory networks in plants and contributes to decipher the transcriptional regulatory code.
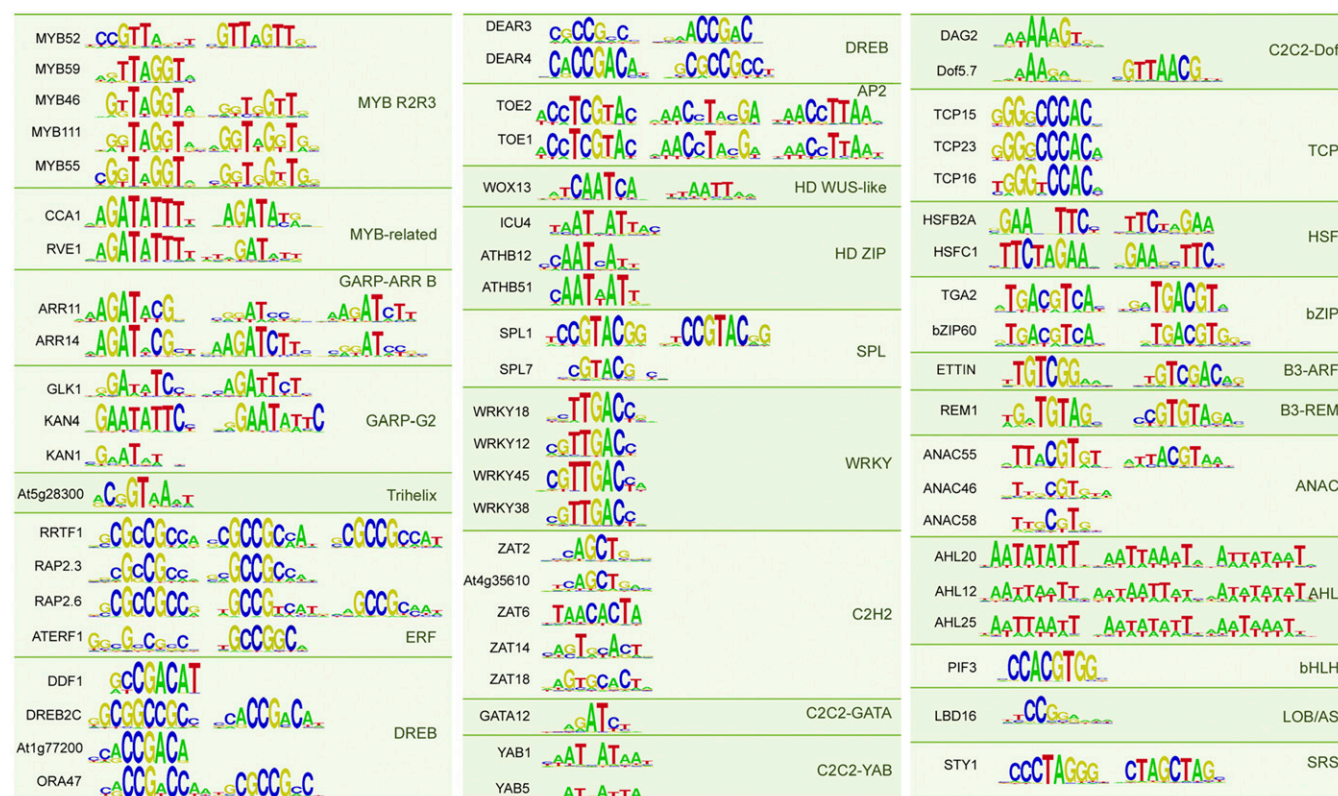

PLANT BIOLOGY

Fig. 1. DNA-binding specificity of plant TFs. Position weight matrix (PWM) representation of top-scoring 8-mers for the TFs indicated. Secondary and tertiary motifs were considered when they differed substantially from primary ones among the lists of top-scoring motifs or after reranking the motifs (*Methods*). TFs are grouped into families or subfamilies according to a previous classification (1).

S2; see also *SI Appendix*, Text S1 and Figs. S1–S27 for a detailed description of the proteins and sequences determined).

Overall, for TFs for which there is information our data are consistent with available data on binding specificity of corresponding families, although some remarkable differences were appreciated. The DEHYDRATION RESPONSE ELEMENT BINDING (DREB) family members DREB2c and DEAR3 (DREB and EAR motif protein 3) recognized a GCC-like element (GCCGCC) with similar affinity than the expected DRE (RCCGAC; R: A or G) for this family (ref. 11 and Fig. 1; see below). DNA motif for WUSCHEL (WUS)-related homeobox WOX13 is partially compatible to that described for the related WUS (ref. 12; TTAATSS; S: G or C). GATA12 recognized the palindromic motif AGATCT, nearly identical to the consensus motif described for this family (13) but differing at a critical residue (WGATAR; W: A or T). The class I TCP16 yielded a binding motif slightly different from the other class I TCPs analyzed, TCP15 and TCP23 (Fig. 1), but matching a class II TCP-motif, similarly to that observed (14, 15).

The Auxin Response Factor (ARF) ETTIN recognized the DNA sequence TGTCGG, partially coincident with the canonical AuxRE (ref. 16; TGTCTC) but differing at the 3′-terminal dinucleotide. However, binding affinity of ETTIN to the AuxRE was notably lower than to the motif that we obtained (*SI Appendix*, Text S1), suggesting that DNA-binding specificity of the ARF family may be broader than initially suspected. The LOB domain protein LBD16 specifically recognized the palindromic sequence TCCGGA, partially differing from the core sequence described for other member of this family (ref. 17; *SI Appendix*, Text S1). Finally, the short internodes/stylish (SHI/STY) family member STY1 recognized a palindromic DNA sequence containing the core CTAG, differing from that proposed for this TF (18).

More interestingly, we determined DNA-binding specificity for several TFs belonging to classes for which there was little or no previous information. The GARP (GOLDEN2,ARR-B,Psr1) members KAN4 and KAN1 recognized similar sequences (Fig. 1), compatible with the motif proposed for KAN1 (19). GLK1 recognized a DNA element containing the palindromic sequence RGATATCY (Y: C or T), compatible to DNA sequences recognized by other GARP or Myb(Myeloblastosis)-related TFs (Fig. 1). TOE1 and TOE2, together with AP2, SMZ, and SNZ, belong to the same phylogenetic clade and they act redundantly in the repression of flowering (7, 20). Both proteins recognized similar sequences, representing a consensus DNA motif for this group of TFs (Fig. 1). Studies on the molecular mechanisms of the function of YABBY (YAB) members are controversial, because these proteins have been proposed to bind DNA either specifically or nonspecifically (21, 22). We analyzed YAB1 and YAB5 and obtained similar binding specificity to A/T-rich elements (Fig. 1), helping us to define a consensus binding-motif for this family as WATNATW. Similarly, the only protein studied so far from the REM class of B3 superfamily is VRN1, which interacts with DNA in a nonspecific manner (23). We obtained a DNA motif for a member of this family (REM1; ref. 24) containing the core sequence TGTAG, representing a cognate sequence for B3 proteins that differs from the consensus binding motifs described for other groups belonging to the same superfamily (Fig. 1).

Previous work to determine DNA-binding specificities of mouse TFs suggested that even closely related TFs have distinct DNA-binding profiles (3). Thus, although proteins with up to 67% amino acid sequence identity may share similar high-affinity binding sequences, they prefer different low-affinity sites (3). To determine whether our data allow differentiating specific DNA-binding patterns among closely related TFs, we analyzed with

more detail DNA-binding specificities of members from two families of TFs, MYB and AP2/EREBP. As expected, we found a strong correlation between amino acid similarity and DNA-binding sequence preference. Structural subfamilies were clearly differentiated, because proteins belonging to the same subfamily showed similar DNA-recognition patterns, reproducing amino acid identities among the members of the same family (*SI Appendix*, Figs. S28 and S29). However, we could identify subtle differences in DNA-recognition patterns of proteins showing up to 79% amino acid identity, suggesting that, despite their similarity, different TFs have distinct DNA-binding profiles.

Remarkably, our analyses also revealed that an unexpectedly high number of TFs recognize secondary elements with similar or slightly lower affinities to their primary ones, which may explain the difficulty to derive consensus binding-sites from ChIP-seq data. Thirty-three of 63 proteins bound to DNA elements partial or completely differing from their primary motifs (Fig. 1). In most cases, secondary motifs represented sequence variants of their corresponding primary elements. In the case of MYBs, two DNA-binding motifs have been described (MBSI and MBSII; ref. 25). Interestingly, only MYB52 recognized both MBSI and MBSII, whereas the other MYBs tested only recognized variants of the MBSII (MBSII and MBSIIG; ref. 25). In the case of B type-*Arabidopsis* response regulator (ARRs), secondary motifs found corresponded to DNA elements described for other members of this family (26). Similarly, ERFs(ethylene response factors), SPL1(squamosa promoter binding protein like), bZIPs (basic leucine zipper), ANAC55(Arabidopsis NAM, ATAF and CUC), and REM1(reproductive meristem) recognized secondary DNA elements, some of them previously proposed (27–30), partially differing from their corresponding primary ones. AT-hook containing TFs (AHLs) yielded several A/T-rich motifs with similar affinities, consistent with their role as modifiers of the architecture of DNA through binding to A/T-rich stretches in the minor groove of DNA (30). Both heat shock factors tested (HSFB2A and HSFC1) recognized identical motifs, representing inverted repeats of the trinucleotide GAA. Actually, both DNA elements may be considered as overlapping halves of a longer motif containing three GAA inverted repeats (TTCNNGAANNTTC), as described for several eukaryotic HSFs (31).

Several AP2/EREBP TFs belonging to the DREB subfamily were recognized with high-affinity GCC and DRE motifs (Fig. 1). These results indicate that DNA binding of DREBs may be more complex than initially suspected, where some DREBs posses a broader range of DNA-binding by recognizing DRE- and GCC-related elements, similarly to that observed for TINY (32). Dof5.7 recognized a canonical binding motif containing the core sequence AAAG (33), but evaluation of secondary motifs yielded a different cognate one. To date, no other Dof protein has been shown to recognize this element, and it may represent an alternative DNA target sequence. The fact that approximately half of the TFs recognized secondary motifs, in some cases completely unrelated to the primary element, suggests that recognition of *cis*-regulatory elements by TFs may be much more complex that initially anticipated.

In line with our results, previous studies with mouse TFs showed that approximately half of them also recognized secondary elements, which were annotated into four different classes (3). In our case, we find more realistic a different classification in which the secondary elements fall into one of the following categories: (*i*) Secondary elements involving variations or extensions on the primary motif, including most (22) of the TFs analyzed. (*ii*) Secondary motifs reflecting overlapping halves of longer motifs (HSFB2A and HSFC1). (*iii*) Secondary motifs corresponding to inverted repeats of a core sequence (ETTIN and STY1). (*iv*) Multiple model, as in ref. 3 (GLK1). (*v*) "Truly" secondary elements found for TOE1, TOE2, WOX13, Dof5.7, and AHLs. It is worth noting that this is an arbitrary classification, and even secondary motifs derived from

variations of the primary ones (class *i*) may retain biological relevance on their own with important implications in the recognition patterns of the TFs, as described above for DREB proteins.

Binding Motifs Are Associated with DNase I Hypersensitive Regions. Binding of TFs to regulatory DNA regions triggers displacement of nucleosomes and chromatin remodeling, resulting in DNase I hypersensitivity (34–36). The ENCODE Project includes mapping of DNase I hypersensitivity sites (DHS) from more than 160 human cell types and tissues, providing an almost saturated mapping of regulatory *cis*-elements (34–36). In *A. thaliana*, although the information is scarce, mapping of DHS from leaf and floral tissues has been useful for the identification of *cis*-regulatory elements (37). We used these datasets for mapping the TF-binding motifs obtained in our PBM assays. Overall, we observed a higher density of TF-binding motifs in DHS fragments from both tissues than their corresponding negative controls (Fig. 2*A*). In addition, we observed particular tissue specificities for some motifs. For instance, density of the HD-ZIP protein ICU4-binding motif was higher in the leaf dataset than in the flower one (Fig. 2*B*), consistent with its role in leaf patterning (38). By contrast, binding sites for the similar ATHB51 were more abundant in DHS derived from floral tissues, indicative of the role of this TF in flower development (39). Taken together, these results indicate that binding sequences identified in PBM assays correlate well with genomic regions exposed to DNase I and, thus, with *cis*-regulatory elements. Moreover, this analysis highlights that combination of TF-binding motifs identified in vitro and genome-wide characterization of DHS in multiple tissues and/or experimental conditions may be an alternative for mapping the TF-binding sites in vivo.

DNA-Binding Sites Have Biological Relevance. As a validation of DNA motifs, we searched in databases for transcriptomic assays of mutant or overexpressing genotypes directly involving the TFs under study. We selected the gene sets deregulated in each of
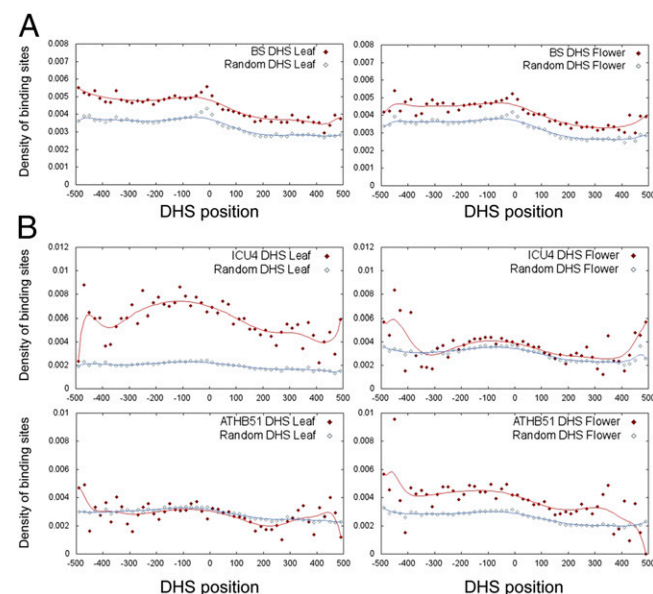
Fig. 2. Binding motifs are associated with DNase I hypersensitive regions. (*A*) Density plots of TF-binding sites in DNase I hypersensitive sites (DHS) from leaf (*Left*) and flower (*Right*) described in ref. 40. Density of TF-binding sites (red diamonds and line) is measured as the number of sites at each position per number of fragments with that length, along DHS fragments (1 kb showed, centered at the middle nucleotide of each DHS). In blue is represented the average density for 100 randomized PWMs. (*B*) Density plots of binding sites recognized by HD-ZIP proteins ICU4 (*Upper*) and ATHB51 (*Lower*).

PLANT BIOLOGY

these experimental conditions and searched at their promoters the DNA elements identified for the TF under study. We observed strong overrepresentation of DNA motifs determined in vitro with sequences in the promoters of deregulated genes in mutant or overexpressing genotypes. In particular, we observed overrepresentation of at least one DNA motif for all of the proteins analyzed by these means (14 in total) and 16 of 20 DNA elements (80%). In some cases (e.g., MYB46 and TGA2), overrepresentation was observed for both primary and secondary motifs obtained in vitro, suggesting their role as *cis*-regulatory sequences (*SI Appendix*, Fig. S30*A*).

We extended this analysis to transcriptomic data indirectly related with TFs tested, either involving structurally similar TFs (*SI Appendix*, Fig. S30*B*), supposed to recognize similar DNA elements, or growth conditions in which the TFs are involved (*SI Appendix*, Fig. S30*C*). In the case of structurally similar TFs, we observed overrepresentation of 30 of 40 (75%) corresponding to 25 different proteins (*SI Appendix*, Fig. S30*B*). It is worth noting that this correlation does not necessarily involve functional redundancy, but rather similarity in high-affinity binding motifs of structurally related TFs. Similarly, 22 of 23 DNA elements

(95.6%) corresponding to 11 TFs were enriched in the promoters of genes differentially expressed in the experimental conditions in which the TFs are involved (*SI Appendix*, Fig. S30*C*).

A further confirmation of the relevance of TF-binding motifs in vivo can be obtained by analyzing enrichment within bound genomic fragments identified by ChIP-seq. In addition to the correlation between in vitro and in vivo data that we observed for PIF5-binding motifs (8), we found a fourfold enrichment of PIF3-binding motifs within ChIP genomic fragments compared with a negative control (*SI Appendix*, Fig. S31), evidencing the good correlation between in vitro and in vivo data.

Validity of Coregulation Data and Binding Motifs to Define Putative Targets. We further evaluated the biological relevance of DNA motifs obtained in vitro in the context of the TF-dependent gene regulation. We hypothesized that, in the case of transcriptional activators, TF-coding genes should have similar expression patterns than their corresponding targets. By contrast, in the case of repressors, TF-coding genes and targets should display opposite expression patterns. We obtained the lists of genes positively and negatively coregulated with the genes encoding the TFs, and
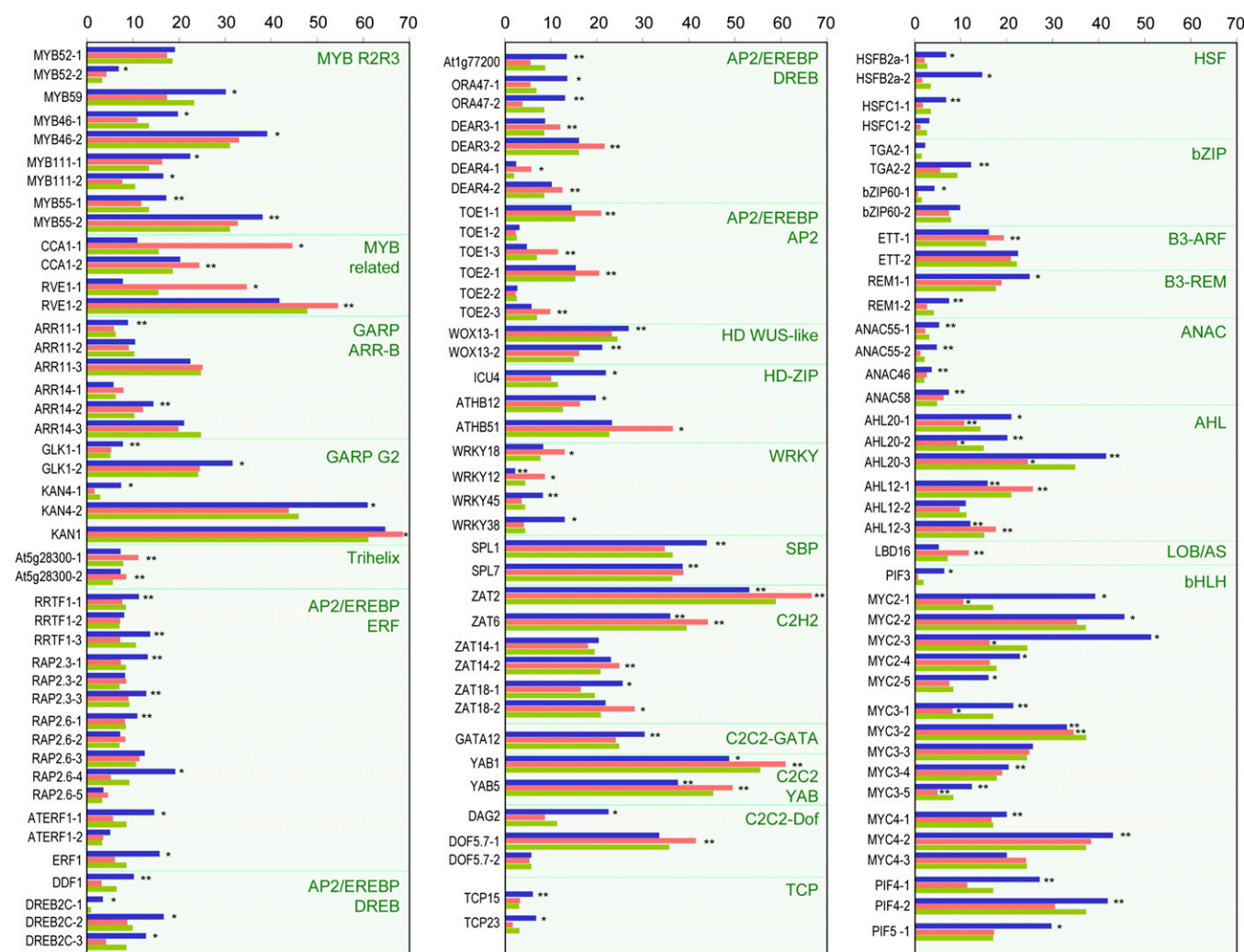
Fig. 3. Evaluation of biological relevance of DNA-motifs from coregulation information. Lists of coregulated genes with the TF-coding genes were scanned for the presence of DNA-motifs obtained in vitro at their promoter regions (1 kb). Frequencies (in %) of genes positively (blue bars) and negatively (red bars) coregulated genes containing the indicated DNA elements are shown. Proportions of genes represented in the ATH1 microarray containing the corresponding elements, and thus representing a random distribution, are represented as green bars. Asterisks indicate the degree of statistical significance in the differences of the proportions indicated as follows: *$P < 0.005$; **$P < 0.05$.

searched at their promoters the DNA elements identified for the TF under study. We also included data relating previously characterized proteins (8, 10, 41). Remarkably, we obtained significant overrepresentation of at least one DNA element for all of the TFs analyzed (69 in total; Fig. 3). When all of the DNA elements were considered, 99 of 122 (81%) were significantly overrepresented (Fig. 3). When considering globally the sets of genes coregulated and deregulated in mutant or overexpressing genotypes, 101 of 122 (82.8%) of the elements had a biological role in the context of gene regulation. These numbers support the extraordinary accuracy of the DNA targets identified in our PBM experiments and underscores the surprisingly high correlation between DNA-binding sequences determined in vitro with *cis*-regulatory elements in vivo. This correlation has been already pointed out in other biological systems (42, 43), and our work shows that it is also the case in plants.

Additional information extracted from this analysis refers to the molecular activity of the TFs studied. Binding motifs for 21 TFs were enriched in the promoters of negatively coregulated genes (Fig. 3), and some of them confirmed in transcriptomic profiles (*SI Appendix*, Fig. S30), suggesting that they act as transcriptional repressors. Of these 21 TFs, 16 were previously proposed to act as repressors, whereas At5g28300, ATHB51, AHL12, LBD16, and Dof5.7 were not (*SI Appendix*, Table S3). We analyzed with more detail the transcriptional properties of these proteins in kinetic assays of luciferase (LUC) activity transiently expressed in *Nicotiana benthamiana* leaves. We cloned the promoter region of one putative target for each TF, selected among negatively coregulated genes containing their corresponding cognate motif. Reporter constructs were coinfiltrated with their corresponding effectors expressing the TF under the control of the constitutive 35S promoter. The results show that four of the five TFs (At5g28300, AHL12, Dof5.7, and LBD16) suppress the expression of their corresponding target promoters, confirming that these four TFs behave as repressors (*SI Appendix*, Fig. S32). In the case of ATHB51, the lack of activity as activator or repressor suggests that we might have failed in the selection of the correct target promoter (*SI Appendix*, Fig. S32). These results confirm the predictive potential of analyzing the sets of coregulated genes for inferring the transcriptional activity of TFs, given that, of 21 predicted, 20 can actually act as repressors.

ChIP experiments have demonstrated that some TFs may recognize DNA sequences at downstream regions of target genes (7–9). We then evaluated the presence of DNA motifs in vitro in downstream regions of coregulated genes. Overall, sequence enrichment was lower than in upstream genomic regions, and observed significant enrichment for 52 of 122 motifs (42.6%), corresponding to 41 of 69 TFs (59.4%) (*SI Appendix*, Fig. S33). When considering the genes coregulated and deregulated in mutant or overexpressing genotypes, we obtained 56 of 122 motifs (45.9%) corresponding to 45 of 69 TFs (65.2%) significantly enriched at downstream regions (*SI Appendix*, Figs. S33 and S34). This enrichment seems to be more conspicuous in the case of transcriptional repressors. Of 21 putative repressors recognizing 36 sites inferred from the analysis of the promoter regions, 25 motifs (69.4%) corresponding to 18 TFs (85.7%) showed significant enrichment at downstream regions of putative targets (*SI Appendix*, Fig. S35). This finding suggests that spatial restrictions for repressors are less strict than for activators, where binding to promoters is preferred.

We next mapped DNA-binding motifs along upstream and downstream regions (1 kb) of coregulated genes. We observed a higher frequency of DNA-binding sites corresponding to transcriptional activators in the promoters of genes positively coregulated near the transcription start site in relation to negatively coregulated or the overall distribution of binding sites (Fig. 4). An analogous distribution was obtained for binding sites corresponding to transcriptional repressors, but relating in this case to negatively coregulated genes (Fig. 4). Consistent with
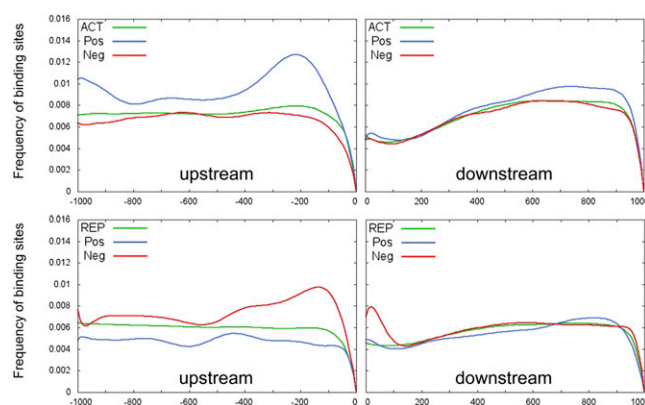
Fig. 4. Mapping of DNA-motifs along upstream and downstream regions. Frequency plots of the average distribution of binding sites at upstream and downstream regions (1 kb) of the genes represented in the ATH1 microarray (green lines), positively coregulated (blues lines), and negatively coregulated (red lines). "0" represents the transcriptional start site in upstream plots, and to stop codon in downstream ones. Plots on the top correspond to binding sites for activators and bottom plots to transcriptional repressors.

our previous analysis, a higher frequency of binding sites was also observed at downstream regions of coregulated genes, although at a lower degree than in promoters (Fig. 4). However, we observed some differences between transcriptional activators and repressors. Whereas binding sites corresponding to activators were enriched at downstream regions 500 bp and further the translation stop codon, binding sites for repressors were particularly enriched near the stop codon, most likely laying at 3′-UTRs (Fig. 4). These results indicate that TF-binding sites corresponding to repressors may be affecting the transcription of target genes, rather than acting as *cis*-regulatory elements. Mapping genome locations associated to transcriptional corepressor complexes will help to unravel the molecular mechanisms underlying transcriptional repression.

Data presented above indicate that the genes coregulated with TF-coding genes are enriched in their corresponding targets, identified as subsets of genes containing cognate motifs at their promoters. To check whether these subsets are functionally related with the TF under study, we analyzed the lists of Gene Ontology (GO) terms significantly enriched in the list of coregulated genes (CRG) and in the subsets of coregulated genes with DNA element (CRGE) for each TF (*SI Appendix*, Fig. S36). We observed a higher enrichment in GO terms related with the function of the TF in CRGE, indicating that these subsets containing cognate DNA motifs at their promoters may represent putative target genes of the TFs.

In addition to help identifying the gene targets of a TF, our work also provides a way to explore how a particular set of coexpressed genes is regulated by scanning their promoters with PBM matrices and identifying the TFs responsible for their regulation among those coexpressed that recognize these sequences. Thus, in summary, the description of TF-binding motifs in vitro and evaluation of their potential as *cis*-elements represent a quick strategy, complementary to ChIP-based methodologies, to define regulatory networks and to help unravel the transcriptional regulatory code.

Materials and Methods

A detailed description of the different methods can be found in *SI Appendix, Text S2*.

Bacterial Expression of Tagged-TFs. cDNAs corresponding to full-length TFs were transferred to destination vector pDEST-TH1 (44), yielding MBP N-terminal fusions. MBP–TF constructs were transformed into BL-21 strain for expression and cultures routinely induced at 25 °C for 6 h with 1 mM isopropyl β-D-1-thiogalactopyranoside.

Identification of TF-Binding Motifs Using PBMs. Recombinant protein extracts obtained from 25 mL of induced *E. coli* cultures and incubations of DNA microarrays were performed as in ref. 11. Normalization of probe intensities and calculation of E-scores of all of the possible 8-mers were carried out with the PBM Analysis Suite (45). We selected the top-scoring 8-mer as the primary DNA element recognized by the TF. Secondary or tertiary DNA motifs were selected if they appeared among top 15 top-scoring motifs and if they differed substantially from their corresponding primary ones. A systematic search for secondary motifs was further carried out by running the Rerank program in the PBM Analysis Suite (45). A list of DNA motifs with E- and Z-scores, their rank positions, PWMs, and microarray designs used is in *SI Appendix*, Table S2.

Analysis of TF-Coding Genes Coexpression Data. Genes positively and negatively coregulated with TF-coding genes were obtained from Genevestigator (46). DNA motifs in the promoter and downstream regions of genes were identified with the Patmatch tool in TAIR. Statistical overrepresentation of DNA motifs was evaluated following a hypergeometric distribution. Analysis of Gene Ontology (GO) terms was performed on GeneCodis (47).

Differential Gene Expression of TF-Involving Genotypes. Gene expression experiments were identified from the literature of from public repositories (GEO, ArrayExpress). A list of gene expression experiments analyzed can be found in *SI Appendix*, Table S4.

Mapping of TF-Binding Matrices. Genomic DHS from leaf and flower tissues described in ref. 37 were scanned for the TF-binding and shuffled matrices in RSAT, and site positions were transformed to density histograms. Upstream and downstream regions of genes were scanned using the same parameters, and frequency histograms were obtained for each binding site. Average frequency histograms for all of the TF-binding sites were obtained and plotted. Plots were obtained with GNUPLOT 4.6.

1. Riechmann JL, et al. (2000) Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* 290(5499):2105–2110.
2. Harbison CT, et al. (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* 431(7004):99–104.
3. Badis G, et al. (2009) Diversity and complexity in DNA recognition by transcription factors. *Science* 324(5935):1720–1723.
4. Jolma A, et al. (2013) DNA-binding specificities of human transcription factors. *Cell* 152(1-2):327–339.
5. Tanay A (2006) Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res* 16(8):962–972.
6. Li XY, et al. (2008) Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. *PLoS Biol* 6(2):e27.
7. Yant L, et al. (2010) Orchestration of the floral transition and floral development in Arabidopsis by the bifunctional transcription factor APETALA2. *Plant Cell* 22(7):2156–2170.
8. Hornitschek P, et al. (2012) Phytochrome interacting factors 4 and 5 control seedling growth in changing light conditions by directly controlling auxin signaling. *Plant J* 71(5):699–711.
9. Huang W, et al. (2012) Mapping the core of the Arabidopsis circadian clock defines the network structure of the oscillator. *Science* 336(6077):75–79.
10. Godoy M, et al. (2011) Improved protein-binding microarrays for the identification of DNA-binding specificities of transcription factors. *Plant J* 66(4):700–711.
11. Shinozaki K, Yamaguchi-Shinozaki K (2000) Molecular responses to dehydration and low temperature: Differences and cross-talk between two stress signaling pathways. *Curr Opin Plant Biol* 3(3):217–223.
12. Lohmann JU, et al. (2001) A molecular link between stem cell regulation and floral patterning in Arabidopsis. *Cell* 105(6):793–803.
13. Lowry JA, Atchley WR (2000) Molecular evolution of the GATA family of transcription factors: Conservation within the DNA-binding domain. *J Mol Evol* 50(2):103–115.
14. Kosugi S, Ohashi Y (2002) DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J* 30(3):337–348.
15. Viola IL, Reinheimer R, Ripoll R, Manassero NG, Gonzalez DH (2012) Determinants of the DNA binding specificity of class I and class II TCP transcription factors. *J Biol Chem* 287(1):347–356.
16. Ulmasov T, Hagen G, Guilfoyle TJ (1997) ARF1, a transcription factor that binds to auxin response elements. *Science* 276(5320):1865–1868.
17. Husbands A, Bell EM, Shuai B, Smith HM, Springer PS (2007) LATERAL ORGAN BOUNDARIES defines a new family of DNA-binding transcription factors and can interact with specific bHLH proteins. *Nucleic Acids Res* 35(19):6663–6671.
18. Eklund DM, et al. (2010) The Arabidopsis thaliana STYLISH1 protein acts as a transcriptional activator regulating auxin biosynthesis. *Plant Cell* 22(2):349–363.
19. Wu G, et al. (2008) KANADI1 regulates adaxial-abaxial polarity in Arabidopsis by directly repressing the transcription of ASYMMETRIC LEAVES2. *Proc Natl Acad Sci USA* 105(42):16392–16397.
20. Jung JH, et al. (2007) The GIGANTEA-regulated microRNA172 mediates photoperiodic flowering independent of CONSTANS in Arabidopsis. *Plant Cell* 19(9):2736–2748.
21. Kanaya E, Nakajima N, Okada K (2002) Non-sequence-specific DNA binding by the FILAMENTOUS FLOWER protein from Arabidopsis thaliana is reduced by EDTA. *J Biol Chem* 277(14):11957–11964.
22. Dai M, et al. (2007) The rice YABBY1 gene is involved in the feedback regulation of gibberellin metabolism. *Plant Physiol* 144(1):121–133.
23. Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C (2002) Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. *Science* 297(5579):243–246.
24. Franco-Zorrilla JM, et al. (2002) AtREM1, a member of a new family of B3 domain-containing genes, is preferentially expressed in reproductive meristems. *Plant Physiol* 128(2):418–427.
25. Solano R, Fuertes A, Sánchez-Pulido L, Valencia A, Paz-Ares J (1997) A single residue substitution causes a switch from the dual DNA binding specificity of plant transcription factor MYB.Ph3 to the animal c-MYB specificity. *J Biol Chem* 272(5):2889–2895.
26. Hosoda K, et al. (2002) Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the Arabidopsis response regulators. *Plant Cell* 14(9):2015–2029.
27. Ohme-Takagi M, Shinshi H (1995) Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell* 7(2):173–182.
28. Jakoby M, et al.; bZIP Research Group (2002) bZIP transcription factors in Arabidopsis. *Trends Plant Sci* 7(3):106–111.
29. Birkenbihl RP, Jach G, Saedler H, Huijser P (2005) Functional dissection of the plant-specific SBP-domain: Overlap of the DNA-binding and nuclear localization domains. *J Mol Biol* 352(3):585–596.
30. Fujimoto S, et al. (2004) Identification of a novel plant MAR DNA binding protein localized on chromosomal surfaces. *Plant Mol Biol* 56(2):225–239.
31. Perisic O, Xiao H, Lis JT (1989) Stable binding of Drosophila heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp recognition unit. *Cell* 59(5):797–806.
32. Sun S, et al. (2008) TINY, a dehydration-responsive element (DRE)-binding protein-like transcription factor connecting the DRE- and ethylene-responsive element-mediated signaling pathways in Arabidopsis. *J Biol Chem* 283(10):6261–6271.
33. Yanagisawa S (2004) Dof domain proteins: Plant-specific transcription factors associated with diverse phenomena unique to plants. *Plant Cell Physiol* 45(4):386–391.
34. Thurman RE, et al. (2012) The accessible chromatin landscape of the human genome. *Nature* 489(7414):75–82.
35. Neph S, et al. (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489(7414):83–90.
36. Bernstein BE, et al.; ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
37. 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(7):2719–2731.
38. Green KA, Prigge MJ, Katzman RB, Clark SE (2005) CORONA, a member of the class III homeodomain leucine zipper gene family in Arabidopsis, regulates stem cell specification and organogenesis. *Plant Cell* 17(3):691–704.
39. Saddic LA, et al. (2006) The LEAFY target LMI1 is a meristem identity regulator and acts together with LEAFY to regulate expression of CAULIFLOWER. *Development* 133(9):1673–1682.
40. Zhang Y, et al. (2013) A quartet of PIF bHLH factors provides a transcriptionally centered signaling hub that regulates seedling morphogenesis through differential expression-patterning of shared target genes in Arabidopsis. *PLoS Genet* 9(1):e1003244.
41. Fernández-Calvo P, et al. (2011) The Arabidopsis bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant Cell* 23(2):701–715.
42. Wei GH, et al. (2010) Genome-wide analysis of ETS-family DNA-binding *in vitro* and *in vivo*. *EMBO J* 29(13):2147–2160.
43. Gordân R, et al. (2011) Curated collection of yeast transcription factor DNA binding specificity data reveals novel structural and gene regulatory insights. *Genome Biol* 12(12):R125.
44. Hammarström M, Hellgren N, van Den Berg S, Berglund H, Härd T (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci* 11(2):313–321.
45. Berger MF, Bulyk ML (2009) Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat Protoc* 4(3):393–411.
46. Hruz T, et al. (2008) Genevestigator v3: A reference expression database for the meta-analysis of transcriptomes. *Adv Bioinforma* 2008:420747.
47. Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A (2012) GeneCodis3: A non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acids Res* 40(Web Server issue):W478-83.

Franco-Zorrilla et al.