A transcription factor hierarchy defines an environmental stress response network

Liang Song, Shao-shan Carol Huang, Aaron Wise, Rosa Castanon, Joseph R. Nery, Huaming Chen, Marina Watanabe, Jerushah Thomas, Ziv Bar-Joseph, Joseph R. Ecker*

INTRODUCTION: Transcription factors (TFs) are often studied one by one or in clusters of a few related factors. However, the integration and networks of transcriptional changes to response to environmental stresses often involve many related TFs. In many organisms, such as plants, overlapping functions can make it difficult to understand how a biologically relevant end result can be achieved via the complex signaling networks controlled by these TFs. To better understand how the reference plant Arabidopsis deals with the stresses incurred by water limitation via the hormone abscisic acid (ABA), we characterized all DNA sequences that bind to the 21 ABA-related TFs in vivo.

RATIONALE: There have been limited systematic studies of stress-responsive TF networks in multicellular organisms. We chose ABA, an essential plant hormone that is required for both development and responses to osmotic stress, as an elicitor to investigate complex gene regulatory networks under stress. Combining differential binding (DB) of 21 ABA-related TFs at a single time point measured by chromatin immunoprecipitation sequencing (ChIP-seq) with differentially expressed genes from a time-series RNA sequencing (RNA-seq) data set, we analyzed the relationship between DB of TFs and differential expression (DE) of target genes, the determinants of DB, and the combinatorial effects of multi-TF binding. These data sets also provide a framework to construct an ABA TF network and to predict genes and cis-regulatory elements important to ABA responses and related environmental stresses.

RESULTS: We found that, in general, DNA binding is correlated with transcript and protein levels of TFs. Most TFs in our study are induced by ABA and gain binding sites (termed “peaks”) after the hormone treatment. ABA also increases the binding of the TFs at most peaks. However, in some peaks, TF binding may be static or even decrease after ABA exposure, revealing the complexity of locus-specific gene regulation. De novo motif discovery enabled us to identify distinct, primary motifs often centrally localized in the ChIP-seq peaks for most TFs. However, it is not uncommon to find motifs, such as the G-box, that are shared by peaks from multiple TFs and may contribute to binding dynamics at these sites. DB of multiple TFs is a robust predictor of both the DE and ABA-related functions of the target genes. Using the DB and DE data, we constructed a network of TFs and canonical ABA pathway genes and demonstrated a regulatory hierarchy of TFs and extensive feedback of ABA responses. On the basis of a “guilt-by-association” paradigm, we further predicted genes whose functions were previously not linked to ABA responses, and we thus functionally characterized a new family of transcriptional regulators.

CONCLUSION: These data sets will provide the plant community with a roadmap of ABA-elicited transcriptional regulation by 21 ABA-related TFs. We propose that dynamic, multi-TF binding could be a criterion for prioritizing the characterization of TF binding events, cis-regulatory elements, and functionally unknown genes in both plants and other species. In our proof-of-principle experiments, ectopic expression of the transcriptional regulators ranked highly in our model results in altered sensitivity to both ABA and high salinity. Together with the fact that our modeling recovered genes related to seed development and osmotic stresses, we believe such predictions are likely applicable to a broad range of developmental stages and osmotic stresses.

The list of author affiliations is available in the full article online.

*Corresponding author. Email: ecker@salk.edu

Cite the article as: L. Song et al., Science 354, aag1550 (2016). DOI: 10.1126/science.aag1550

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aag1550

Downloaded from http://science.sciencemag.org/ on July 9, 2021

science.aag1550
A transcription factor hierarchy defines an environmental stress response network

Liang Song,1 Shao-shan Carol Huang,1 Aaron Wise,2 Rosa Castanon,3 Joseph R. Nery,4 Huaming Chen,3 Marina Watanabe,1 Jerushah Thomas,1 Ziv Bar-Joseph,2 Joseph R. Ecker1,3,4*

Environmental stresses are universally encountered by microbes, plants, and animals. Yet systematic studies of stress-responsive transcription factor (TF) networks in multicellular organisms have been limited. The phytohormone abscisic acid (ABA) influences the expression of thousands of genes, allowing us to characterize complex stress-responsive regulatory networks. Using chromatin immunoprecipitation sequencing, we identified genome-wide targets of 21 ABA-related TFs to construct a comprehensive regulatory network in *Arabidopsis thaliana*. Determinants of dynamic TF binding and a hierarchy among TFs were defined, illuminating the relationship between differential gene expression patterns and ABA pathway feedback regulation. By extrapolating regulatory characteristics of observed canonical ABA pathway components, we identified a new family of transcriptional regulators modulating ABA and salt responsiveness and demonstrated their utility to modulate plant resilience to osmotic stress.

Abscisic acid plays a pivotal role in optimizing water use in plants and is required for both seed development and responses to multiple environmental stresses, such as drought and high salinity. In *A. thaliana*, ABA is recognized by the PYR/PYL/RCAR receptor proteins (6–8). Binding of ABA triggers the interaction of PYR/PYL/RCARs with group A PP2C protein phosphatases and derepresses the SnRK2 protein kinases (7, 9). SnRK2s subsequently activate substrates such as TFs and elicit ABA responses (10, 11). Whereas many TFs are known to contribute to the ABA responses (12), little is known about their target genes and the way these targets are combinatorially regulated. In vitro approaches, including the recently described *Arabidopsis* cistrome data set, have enabled identification of DNA motifs and inference of the associated TFs (13–17). However, accurate predictions are still challenging due to many large, multi-member TF families. Furthermore, it is difficult to establish a direct link between TF binding and transcriptional changes or to address the dynamics of TF regulation through in vitro assays. Therefore, we applied ChIP-seq to unambiguously discover TF targets, examine stimulus-influenced TF binding dynamics, and link these findings to subsequent transcriptome changes.

ChIP-seq analyses of ABA responses

We profiled the genome-wide binding dynamics of a diverse collection of TFs using ChIP-seq to develop an in planta ABA transcriptional regulatory network. We first surveyed ABA-responsive transcripts in *A. thaliana* by generating strand-specific RNA-seq libraries from 3-day-old etiolated whole seedlings treated with either 10 μM (±)-ABA or an ethanol-containing mock for 1, 4, 8, 12, 24, 36, and 60 hours (Fig. 1A). Among 18,310 expressed genes, 3061 are differentially expressed [false discovery rate (FDR) < 0.01 (table S1) (18)] for at least one time point. One hour of ABA treatment leads to moderate DE of many genes, and most transcriptional responses plateau after 8 hours (Fig. 1B and fig. S1). On the basis of the gene expression data, we performed ChIP-seq experiments 4 hours after dosing with ABA. We selected TFs based on responsiveness to ABA and published evidence, aiming to provide a comprehensive representation of TF families (fig. S2 and tables S2 and S3). In general, highly expressed and responsive TFs were chosen in each representative TF family because in the context of an in planta experimental framework, the impact of these TFs on gene expression can be more effectively investigated compared with that of their weakly expressed homologs. All TF genes were epitope-tagged by a recombinering-based approach (19), mostly with large DNA transformable artificial chromosomes, allowing the TFs to be expressed under their native promoters and genomic context (table S3). The final data set consisted of 122 ChIP-seq experiments involving 21 TFs from 11 families, including mock- and ABA-treated conditions.

Overall, the number of binding sites (“peaks”) varies greatly across TFs and between treatments (Fig. 1C). Most TFs gain binding sites across the genome after ABA treatment (Fig. 1C), consistent with the fact that these TFs are induced by ABA at both the transcript and protein level (Fig. 1D and figs. S2 and S3). As exemplified by CYP707A1 and HAI2 (20, 21), two important genes regulating ABA catabolism and signaling, respectively, the dynamic binding of TFs elicited by ABA is often accompanied by altered transcript abundance of the target genes (Fig. 1D). Accounting for binding location and strength, a comparison of the genome-wide binding profiles of these TFs revealed that the TFs are generally grouped by family and known physical interactions (Fig. 1E) (22). The binding profiles between NAM, ATAF, and CUC (NAC) and other TF families become more similar after ABA treatment (Fig. 1E, box A versus M), which suggests that ABA prompts coordinated regulation of target genes by these TFs.

Hormonal effects on TF binding and expression of target genes

We observed substantial changes in TF binding at promoter regions of several known components of the ABA signaling pathway (Fig. 1D), so we investigated whether dynamic binding may predict gene function in the ABA pathway. To quantify hormone-dependent, locus-specific changes of TF binding, we compared ChIP-seq peaks of each TF between ABA- and mock-treated conditions by performing differential binding analysis of the sequencing reads under the peaks (23). Three measures of differential binding were calculated for each peak: (i) normalized read count change (RCC) that measures absolute changes of binding, (ii) fold change (FC) that measures relative changes of binding, and (iii) statistically significant differential binding (FDR). Because there are limited
down-regulated binding events in our data set, we focused on up-regulated binding to determine the optimal cutoff of RCC, FC, and FDR to define the relationship between dynamic binding targets and genes involved in the ABA response. We extracted three groups of *A. thaliana* genes (table S4) on the basis of gene ontology (GO) annotation (24). Group 1 contains 493 genes involved in the ABA response. Group 2 contains 1452 genes involved in responses to either ABA or other related processes such as water deprivation, osmotic stress, salt stress, cold, seed development, and stomatal movement. Group 3 contains 999 genes involved in responses to other hormones after excluding genes shared with group 2. Three observations emerged from comparing these lists to the TF target gene lists defined by various thresholds on RCC, FC, and FDR. First, when group 1 and 2 genes are used as a reference set, the percentage of TF targets overlapping with the set increases with the number of bound TFs (Fig. 2A, left and middle panels). By contrast, there is very little, if any, increase when group 3 (other hormone genes) is used as the reference (Fig. 2A, right panel). Second, an increase of RCC and FC threshold beyond the top 20% boosted the percentage of target genes involved in ABA-related responses but not the percentage of genes related to other hormones (Fig. 2A, right panel). This improvement is even more obvious for genes targeted by multiple TFs. Last, FDR thresholds of 0.1 and 0.2 show few differences across all analyses. These results support the premise that dynamic binding by multiple TFs is an important

---

**Fig. 1.** TF identity and hormone treatment determine genome-wide binding profiles. (A) Growing *A. thaliana* in hydroponics allows convenient buffer exchange for hormone treatment. (B) DREM-reconstructed RNA expression paths 60 hours after exposure to ABA. Each path corresponds to a set of genes that are coexpressed. Split nodes (green diamonds) represent a temporal event where a group of genes coexpressed up to that point diverges in expression, most likely due to regulatory events. Most splits are observed up to and including the 4-hour time point, indicating that the majority of regulatory events occur during the first 4 hours. (C) The number of ChIP-seq peaks varies greatly between TFs and treatment conditions. (D) ABA mediated differential gene expression and altered dynamics of TF binding, as exemplified by the *CYP707A1* and *HAI2* genes. (E) Comparison of binding correlations based on normalized ChIP-seq read counts near binding sites shows that TFs from the same family often have similar binding profiles. TF-TF interaction (bZIPs and NF-Y, black dashed box) and hormone treatment (RD26 and ANAC032, dotted boxes A and M for ABA- and mock-treatment, respectively) also contribute to binding profile similarities between TFs.
feature to specifically recover genes involved in ABA-related responses. We thus chose the top 20% of RCCs and FCs, as well as FDR = 0.1, as the cutoffs for follow-up analyses. As shown in Fig. 2B and fig. S4, peaks passing all three thresholds were designated as top-ranked up-regulated (“top up”) or down-regulated (“top down”), whereas those failing all thresholds were designated as static; all remaining peaks were classified as moderately up-regulated (“moderate up”) or down-regulated (“moderate down”). For all tested TFs except FBH3 and HB5, peaks tend to gain binding instead of maintaining or losing binding after ABA treatment.

We then explored the relationship between dynamic TF binding triggered by ABA treatment and gene expression. The Dynamic Regulatory Event Miner (DREM) (25) reports 11 paths of DE genes for the first 8 hours of ABA treatment (Fig. 2C). As shown in fig. S5, combining DREM with DNA motifs from PBM (protein binding microarrays), AGRI (Arabidopsis Gene Regulatory Information Server), and DAP-seq (DNA affinity purification sequencing) databases recovered few TFs in our data set, likely because of a low overlap of these TFs with the databases (13, 15, 26). The DREM model identified TFs from all chromatin immunoprecipitated families except for CCAAT-HAP3 and CCAAT-HAP5, which do not bind DNA as a monomer in in vitro assays (27).

In addition, although TF binding was examined at a single time point, we observed a positive correlation between the number of dynamically bound TFs and the magnitude of DE across all time points (fig. S6), which suggests that TF binding data at 4 hours after dosing with ABA can explain a broad temporal span of gene expression. ABA-related GO terms such as “seed development” and “response to salt/osmotic stress/water deprivation” were enriched in up-regulated genes, whereas a few growth-related terms such as “response to auxin stimulus” and “cell wall organization” were enriched in down-regulated genes (Fig. 2D). We observed a distinct distribution of dynamic binding category across DREM paths (Fig. 2E, fig. S1, and table S5). The extent of multi-TF dynamic binding is associated with the magnitude of differential gene expression. For example, highly up-regulated genes are often targeted by multiple TFs through top-up binding. Moderately up-regulated genes are more commonly targeted by multiple TFs through moderate-up binding. Down-regulated genes are rarely associated with top-up binding. Instead, these genes are predominantly associated with either static binding by multiple TFs or down-regulated TF binding. These data suggest that DE at the whole-seedling level is often subject to a combinatorial regulation by multiple TFs. As an independent validation, we built a regression model of DE using peak signals in ABA- and mock-treated conditions as features without hard thresholds for the level of dynamic binding. The resulting model reveals that multiple TF binding features such as ZAT6, NF-YB2, and ABF factors in both ABA- and mock-treated conditions contribute to DE of target genes (Fig. 2F and fig. S7).

**Determinants of differential TF binding**

With the discovery of tens of thousands of differential binding events, we were interested in whether we could identify features that may predict...
binding dynamics. First, we performed motif discovery by MEME-ChIP (28) to identify enriched motifs of all 21 chromatin immunoprecipitated TFs from the strongest 600 peaks after either ABA or mock hormone treatment. A complete collection of the motifs is available at www.ABAtf.net. To investigate whether there are additional motifs that correlated with TF binding dynamics, we also performed motif discovery on both dynamic and static peaks for a handful of TFs. These factors (NF-YB2, ABF1, ABF4, FBH3, MYB3, RD26, ZAT6, and HB7) were selected to represent a variety of TF families. Pairwise comparison of primary and secondary motifs discovered from dynamic and static peaks across the selected TFs revealed three major clusters (Fig. 3A). Cluster 1 motifs are composed of \((AG)_n\) repeats. Cluster 2 motifs contain a \((A/G)(A/C)CC(A/C)\) consensus sequence, whereas cluster 3 comprises G-box motifs (Fig. 3B). To examine the contributors to binding dynamics, we used linear regression to model the FC of binding as a function of variables, including basal binding of the TF (under mock treatment) and the number of occurrences (counts) of a set of non-redundant sequence features that capture the diversity of motifs. These sequence features were selected from major clusters of all motifs found in the strongest 600 peaks (motifs of clusters A to I) and the dynamic and static peaks (motifs of clusters 1 to 3) (fig. S8). Examination of the regression coefficient \(P\) values (Fig. 3B) suggests that the primary motifs of ABF (which also represent cluster 3 motifs) and ANAC TFs are associated with enhanced dynamic binding, whereas basal binding and cluster 2 motifs are associated with a negative effect on binding dynamics for a broad range of TFs (Fig. 3B and table S6). Including cluster 3 or cluster 1 motifs in the regression results increases the explained variability by up to 20% (Fig. 3B). To visualize the effect of the cluster 3 G-box motif and the cluster 2 motif at the resolution of individual peaks, we plotted basal binding of TFs quantified by normalized read count against log2 FC of binding after ABA treatment and assigned a color to individual binding events on the basis of the count of motifs in the same peak (Fig. 3C). The proportion of peaks containing the cluster 3 motif increases along with log2

---

**Fig. 3. Determinants of TF binding dynamics.** (A) Hierarchical clustering of motifs enriched in dynamic and static peaks revealed three clusters. Each entry in the distance matrix is \(-\log_2(P\text{ value})\) of motif similarity reported by Tomtom (28). (B) Linear regression of differential binding using basal binding and non-redundant sequence features identified positive and negative determinants of dynamic TF binding. Heatmap colors map to two-tailed \(t\) test \(P\) values on the regression coefficients for the null hypothesis that the coefficient is zero. The sequence features were selected from motifs enriched in the strongest peaks in ABA- and mock-treated conditions, as well as dynamic and static peaks (fig. S8). (C) Scatterplots show basal binding of TFs quantified by normalized read count in the peak (x axis) against log2 FC of TF binding after ABA treatment (y axis), with the color of each dot mapped to the number of indicated motifs in the same peak. The occurrence of cluster 3 and 2 motifs over the distributions of log2 FC of binding is shown in the histograms, with the same color code as the scatter plot. The proportion of peaks containing the cluster 3 motif increases along with log2 FC of TF binding for the indicated TFs, whereas the proportion of peaks containing the cluster 2 motif is negatively correlated with log2 FC of TF binding.
FC of binding, whereas the proportion of peaks containing the cluster 2 common motif is negatively correlated with log2 FC of binding. These data suggest that the binding of a TF to the cluster 3 motif (likely the ABFs) and the binding of an unknown family of TFs to the cluster 2 motif regulate the binding dynamics of neighboring TFs positively and negatively, respectively.

**Construction of an ABA response network**

To confirm that dynamic binding is more robust than total binding in predicting gene expression and genes involved in ABA and related responses, we compared the expression and functional composition of genes grouped by the number of targeting TFs through either any type of binding or top-up binding (Fig. 4A). Representation of both genes associated with ABA-related GO terms and ABA up-regulated genes increases more rapidly with the increasing number of TFs that have top-up binding. Therefore, we decided to use top-ranked dynamic TF binding triggered by ABA treatment to demonstrate the wiring of this ABA network using the core ABA metabolic and signaling genes and to calculate the hierarchical height of TFs in the network (Fig. 4, B and C, and fig. S9A). TFs in the network are organized into three tiers by their hierarchical height (Fig. 4C and fig. S9A). The level of DE of lower-tier TFs is often amplified compared with that of upper-tier TFs, resulting in greater changes in binding dynamics, likely because of greater protein accumulation (figs. S3, S4, and S9A). Negative regulators of ABA response—including genes encoding ABA catabolic enzymes, protein phosphatase 2Cs, and E3 ligases—are often induced by ABA and are heavily targeted by multiple TFs through highly up-regulated TF binding (Fig. 4C). By contrast, positive regulators of the ABA response can either be up-regulated due to increased TF binding or down-regulated due to reduced TF binding (Fig. 4C). These results point to a transcriptional feedback strategy in the ABA response, presumably to allow rapid restoration of normal growth once stress is lifted. Because some transcriptional responses triggered by ABA are similar to those triggered by natural stresses (figs. S9, A and B) such as high salinity conditions, we also expect to see a similar organization of regulatory networks for other osmotic-related stresses.

**Extensive targeting by ABA-responsive TFs**

The expression of genes in one hormone response pathway to another.

No GO term besides the ABA-related ones was enriched in DE genes heavily targeted by the 21 TFs through top-up binding. This is partially because more than one-third (12,336/33,601) of the genes in the A. thaliana genome still have no information regarding their biological processes (BP) (Fig. 5A and table S4). On the basis of a “guilt-by-association” paradigm (30), we speculate that many BP-unknown genes in Fig. 5A are also involved in ABA responses (table S8). As a proof of principle, we functionally characterized a family in which all members are BP-unknown and differentially expressed in response to ABA. In particular, three members in this family (At3g48510, At15g50360, and At15g40790) are heavily targeted by TFs through top-up binding (Fig. 5, A and B). Little is known about this family except that the proper expression of At15g48510 relies on core ABA signaling (32). In addition, predicted proteins of this family contain no known domains. We generated dexamethasone (DEX)-inducible lines expressing green fluorescent protein (GFP) fusion of the two most heavily targeted genes (At15g48510 and At15g50360). Analysis by RNA-seq showed that a few hundred DE genes were consistently identified from both short-term (4 hours) and long-term (10 days) DEX induction of the two genes (Fig. 5C and table S6). To reflect their regulation and function, these genes were named Dynamic Influencer of Gene expression 1 (DIG1) and DIG2, and their homologs were named DIG-likes (DILs). ABA-related GO terms, such as “response to water deprivation,” were enriched in DIG down-regulated genes (Fig. 5C). Confocal imaging further showed that DIGs were localized to the nucleus (Fig. 5D). We then tested whether the DIGs are transcriptional regulators. ChIP-seq of DEX-inducible GFP-DIGs showed that the DIGs bind chromatin. Moreover, stronger binding was observed in the promoter of DIG down-regulated genes than up-regulated ones or non-DE genes (Fig. 5, E and F). De novo

---

**Fig. 4. TF network integrates expression and connectivity features of genes in the ABA response.**

(A and B) Expression and functional composition of all genes (A) and TF genes (B) are grouped by the number of targeting TFs through either any kind of binding or top-up binding. Compared with any type of binding, top-up binding is a better predictor for both ABA-related BP functions and DE. The number of genes in each bin is shown in black. The bins to which the TFs included in this study belong are indicated at the top of (B). (C) ABA pathway genes are subject to extensive feedback regulations and multi-TF dynamic binding. Chromatin immunoprecipitated TFs are arranged in three tiers by normalized hierarchy height. Target genes are grouped by function. Node color depicts changes of transcript abundance after 4 hours of ABA treatment. Edge color corresponds to TF binding dynamic categories.
motif discovery allowed us to identify a CCAAT(n)8 ABRE motif strongly enriched near the DIG1 binding sites within 1 kb of DIG down-regulated genes. By contrast, either a weaker motif or no similar motif was enriched near DIG binding sites in the corresponding regions of non-DE genes or DIG up-regulated genes (Fig. 5, G and H). Several ABA-responsive or developmental TFs are targeted by DIGs and differentially expressed transgenic seedlings shows that DIG1 is nuclear-localized. (E and F) Metagene profiles (E) and heatmaps (F) of normalized ChIP-seq read counts surrounding DIG DE genes. Down-regulated genes are often associated with strong DIG binding in their promoters. (G) Empirical cumulative distributions of \(-\log_{10}(P\text{-value})\) of ChIP-seq peaks of DIG1 show that it binds more strongly to DIG down-regulated genes than to up-regulated or non-DE genes. (H) A CCAAT(n)8 ABRE motif is strongly enriched near DIG1 binding sites residing within 1 kb of DIG down-regulated genes. Either a weaker motif or no similar motif is enriched in the corresponding regions of non-DE genes or DIG up-regulated genes. (I) Induction of DIGs results in DE of ABA- and developmental-related TFs.
upon the induction of DIGs (table S9). Among these, ATA1, HY5, and ABF3 have been linked to ABA sensitivity (33–35), whereas HY5, SCL3, and perhaps IAA9 have developmental roles (36, 37). Sequence analysis revealed that DIGs are conserved between monocots and dicots (fig. S11). A remotely related clade of DIG contains the gene Sdr4, which regulates seed dormancy in rice (38) (fig. S12). The Sdr4 paralog in Arabidopsis is also dynamically targeted by multiple ABA-responsive TFs and differentially expressed in response to ABA (table S1). However, the functionally important amino acid residues of Sdr4 are not conserved in the DIGs and their homologs (fig. S13) (39). Therefore, genes in the DIG and Sdr4 clades may exert ABA-related functions through distinct mechanisms. Finally, inducible expression of DIGs enhances ABA sensitivity, as assayed by cotyledon greening (Fig. 6, A and B) and lateral root growth (Fig. 6D). Similarly, enhanced growth inhibition of DIG lines can also be observed after prolonged growth under high-NaCl conditions (Fig. 6, C, E, and F). Combined, our results suggest that DIGs are a family of transcriptional regulators with broad roles that include regulating gene expression that affects ABA sensitivity and salt stress responses.

Conclusions

We performed a systematic study of a transcriptional network by combining dynamic binding data of 21 TFs and time series RNA-seq data in response to a stimulus by the plant hormone ABA. We found that dynamic TF binding measured at a single time point correlated with the transcriptome changes over a prolonged span of time. Consistent with yeast and animals (2, 3, 39, 40), transcription of genes in Arabidopsis is often subject to a complex regulation of multiple TFs. We further demonstrated that dynamic binding, especially by multiple TFs, is more functionally relevant than static TF binding in correlation with differential gene expression. We speculate that this is because an expression scheme with coordinated changes in the binding dynamics of multiple TFs would ensure robust responsiveness of target genes to a stimulus. This observation would have a broad application to both plants and other species, including prioritizing studies of (i) TF binding events and cis-regulatory elements and (ii) functionally unknown genes in a pathway.

In plants, studies of transcriptional regulation are often focused on master regulators (33, 41, 42). Our data confirmed the importance of master regulators in plants. For instance, we showed that ABFs and a physical interactor (NF-YB2) ranked as top contributors to explain gene expression. In addition, the primary binding motif of ABFs also enhances the binding dynamics of many other ABA-responsive TFs. However, more than just the master regulators are required to attain complex transcriptome changes to a stimulus, as many ABA-responsive genes are dynamically targeted by multiple TFs. Therefore, ABA response can be viewed as orchestrated by a handful of master regulators and facilitated by other TFs, where coordinated signaling and regulatory response lead to a rapid elicitation of transcriptome changes.

As indicated by GO annotation, network analyses of this study recovered genes affecting all aspects of ABA-related processes, such as seed development and response to osmotic stresses. In planta ectopic expression of several members of a newly discovered family of transcriptional
regulators also exhibited altered response to both ABA and high salinity. Therefore, although we carried out the experiments using seedlings, the discoveries may be directly applicable to a broader range of development stages and stress scenarios. Emerging technologies to optimize plant water use have been developed based on the in-depth characterization of ABA perception (45). Knowledge derived from further studies of the genes uncovered in this study may also prove valuable to global agriculture, possibly enabling new strategies for plants to respond to the challenges of ongoing drought and groundwater depletion in changing environments.

Materials and methods

Plant materials

Recombineering lines for the ChIP-seq experiments were generated as previously described (19) with minor modifications. A YPet-6xHis-3xFLAG tag and a 3xFLAG-YPet tag were designed for C. aag1550-8. A YPet-6xHis-3xFLAG tag and a 3xFLAG-YPet tag were designed for C. aag1550-8 with minor modifications. A YPet-6xHis-3xFLAG tag and a 3xFLAG-YPet tag were designed for C.

ChIP-seq experiments and analysis

0.4 g seeds were surface sterilized by 50% bleach + 0.05% Triton-X100 for 10 min. After 4 days of stratification at 4°C, seeds were spread on nylon mesh (Component Supply, FL, USA, cat. # U-CMN-215) in 6 hydroponics (Sigma-Aldrich, MO, USA, cat. # P1552) containing 1x pH buffered LS basal salts. After exposure under light for 4 hours to enhance germination, seeds were grown in dark at 22°C for 3 days. Etiolated seedlings were then switched to 1x LS buffer containing either (-)ABA (MP Biomedicals LLC, CA, USA, cat. # 190673) dissolved in 100% ethanol at a final concentration of 10 μM or ethanol alone as mock and treated for 4 hours in dark before ChIP as previously described (46). Briefly, harvested seedlings were cross-linked by 1% formaldehyde solution (Sigma-Aldrich, cat. # F8775) under vacuum for 20 min. After nuclei isolation, chromatin was sonicated to 100-400 bp fragments. Tagged TFs in the transgenic lines were immunoprecipitated by a rabbit polyclonal anti-GFP antibody (Thermo Fisher Scientific, MA, USA, cat. # A1122). After elution, reverse crosslinking and DNA purification, Illumina TruSeq libraries were constructed according to manufacturer's protocols. All ChIP-seq experiments in both ABA and ethanol mock-treatment conditions were done with biological replicates. Uniquely mapped sequencing reads to the TAIR10 genome assembly (Bowtie v0.12.7) (47, 48) were used to call peaks by the IDR pipeline utilizing MACS2 (49) with mock IP of wild-type Col-0 ChIPped by the anti-GFP antibody as a control. Peaks with a P ≤ 1 × 10^-16 were kept and differentiated binding of TFs were analyzed by DiffBind (v1.10.1 with edgeR v3.0.8) (23). To calculate TF binding similarity in Fig. 1E, the center of peaks (termed “summits”) of all ChIPped TFs were pooled together to create a union list. Sequencing coverage within 50 base pairs of summits in the union list was counted and normalized by deepTools (v1.5.8) (50). Pairwise Pearson correlation between samples was used as entries in the distance matrix to plot the heat map in Fig. 1E. Hierarchy height of ChIPped TFs was calculated as previously described (40)

$$h = (O - I)/(O + I)$$

where O and I are out-degree and in-degree of examined TF through top-ranked dynamic binding. Peaks in each dynamic binding categories were associated to TAIR10 annotated genes within 1000 bp from the summit of the peaks, using the R BioConductor package ChIPpeakAnno (v2.12.1) (51).

Motif discovery and modeling the contribution of individual motifs to TF binding dynamics

De novo motif discovery was carried out by MEME-ChIP (MEME 4.9.1) using a background file calculated from TAIR10 intergenic sequences (47). Top five enriched motifs identified within 50 base pairs of the summits were filtered at e-value cutoff of 1e-05. To model the contribution of individual features, a set of nonredundant sequence features were selected to represent the overall motif diversity. To do this we first assembled a set of 135 motifs in our data set, consisted of the two most enriched motifs in the top 600 peaks in ABA- and mock-treated conditions for each TF, as well as top five motifs enriched in dynamic and static peaks for all TFs. The motifs were clustered by applying hierarchical clustering using motif distances calculated by Pearson correlated coefficients as column comparison metric and Ungapped Smith-Waterman alignment method (32, 53). Dynamic tree cut of the clustering dendrogram (54) identified 19 major clusters (color of dendrogram branch and the left of the annotation tracks in fig. S8). As several of the clusters contain similar motifs (for example, the G-boxes and the AG-rich motifs are split into multiple clusters), we selected 11 sequence features to capture the diversity in this set of motifs identified by dark red color of motif name and dark red color in the right annotation track in fig. S8. Basal binding was measured as log2(normalized read counts) under mock treatment and occurrences of motifs were assessed by FIMO (55) at a P value cutoff of 0.0004. These features were used in the lm() function of R (v3.2.2) to fit log2(fold change) of the binding of indicated TFs in Fig. 2D between ABA- and mock treatment. Relative changes of explained variance was calculated as

$$(R^2 - R_0^2)/R^2$$

where $R^2$ and $R_0^2$ are the adjusted $R^2$ from lm() output that includes and excludes cluster II or cluster III motif as a feature, respectively.

RNA-seq experiments and analysis

For ABA time series experiments, two biological replicates of 3-day-old etiolated, hydropnoc-grown wild-type Col-0 seedlings were treated either by 10 μM (±)-ABA (MP biomedicals, LLC, cat. # 190673) dissolved in ethanol or ethanol-only mock control for 24, 48, 72, 96, 120, 144, and 168 h. For DEX treatment, short-term experiment was carried out by treating 3-day-old etiolated, DEX-inducible GFP-DIG1, GFP-DIG2, or GFP lines with 10 μM DEX (Sigma Aldrich, cat. # D9841); long term experiment was carried out by growing the same lines of plants containing 500 nM DEX for 10 days. Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, CA, USA, cat. # 74903), and cDNA libraries were constructed using the TruSeq Strand-stranded Total RNA LT Sample Prep Kit (Illumina, USA, cat. # 15032611) according to manufacturer’s instructions. Single-end reads were generated by the HiSeq 2500 Sequencing System (Illumina) and mapped to TAIR10 genome assembly using TopHat 2 (v2.0.8) (56). Mapped reads with mapping score equal to or larger than 10 were counted by HTSeq (v0.5.4) (57) and analyzed by edgeR (v3.6.2) (58) to identify differentially expressed genes using contrasts between ABA- and mock-treated samples at each time point and false discovery rate 0.01 or 0.05 as thresholds.

DREM

The Dynamic Regulatory Events Miner (DREM) (25, 58), integrates TF-gene interactions from ChIP-seq experiments with time series gene expression data to identify patterns of temporal gene expression, the associated regulators and the dynamics of the interactions. Splits in the reconstructed network (green nodes in Figs. 1B and 2C and figs. S1 and S5) represent divergence of genes that are co-regulated by ChIP-seq (fig. S1) or inferred from the AGRIS information (the timing of the splits) with the DREM with the TFs that are predicted to regulate them, allowing us to associate the temporal information (the timing of the splits) with the interaction information either directly measured by ChIP-seq (fig. S1) or inferred from the AGRIS database (26), PBM (33), and DAP-seq (55) data (fig. S5). The analysis performed here used the log fold change of 3061 DE genes (table S1) identified in the ABA time series RNA-seq data. DREM paths were created using all DE genes without further filtering.

For GO enrichment in DE genes targeted by categories of dynamic binding peaks (Fig. 2D),
we defined the genes in distinct DREM paths as foreground, and all expressed genes as background, and retrieved the Functional Annotation Chart with EASE score (modified P value) threshold of 0.1 and count threshold of 2, using functional annotations provided by the R BioConductor package RDAVIDWebService (29) to query the DAVID web service (60). The GO terms in GO_TERM_BP_FAT with FDR ≤ 1% from all target gene sets are combined, and the enrichment P values of these terms are retrieved for each gene set to create the heatmap in Fig. 2D. If a term is not reported to be significant for a target set, its P value is set to 0.1 (the P value threshold).

Modeling the contribution of individual TFs to gene expression

We adopted an approach similar to previous regression-based models that relate gene expression to TF binding (67, 62). We first defined TF affinity score (TFAS), $A_{ij}$ for TF $j$ on gene $i$, using the peak closest to the TSS of the gene

$$A_{ij} = g e^{rac{e}{d} - d_{ij}}$$

where $g$ is the log2 TMM normalized read counts of the peak, $d$ is the distance of the peak summit to TSS, $d_{ij}$ is set to 1000. For $N$ genes and $M$ TF, we constructed one $N \times M$ TFAS matrix for the ABA treatment and one for the ethanolic mock treatment, and concatenated these two matrices horizontally to create a final $N \times 2M$ matrix $A$. We centered and scaled each column of $A$ and fit a log-linear model

$$\log Y_i = \sum_{j=1}^{2M} \beta_j A_{ij} + \epsilon_i$$

where $Y_i$ is the fold change in expression of gene $i$ at 4 hours ABA treatment compared to mock. We limited the model training and testing to genes that are differentially expressed at 4 hours with FDR ≤ 0.01 and those that are not differentially expressed at all time points (FDR > 0.7). A glmnet regression model (63) was trained on 75% of the genes by five repeats of 10-fold cross-validation using the caret package in R (66) with tuning metric set to RMSE and the elastic net mixing parameter $\alpha = 0$ to allow selection correlated TFAS features. The “best” rule was used to choose a value for the tuning parameter (in this case, the regularization parameter $\lambda$), i.e., a value that minimized the average RMSE of the regression on the 50 resampling of the training set. The glmnet model was then fitted using the chosen $\lambda$ value to arrive at the regression coefficients in the final model. The unscaled coefficients of the TFAS features are plotted as binding feature importance in Fig. 2F.

Confocal imaging

Nine-day-old DEX inducible GFP and GFP-DIGI seedlings grown on 1x LS plates containing 200 nM DEX and 300 nM ABA were imaged by Zeiss 710 confocal microscope under an Argon laser at 488 nm. GFP signal was captured within the 490- to 548-nm emission window and was pseudo-colored in green. Auto-fluorescence from chloroplasts was captured within the 569- to 695-nm emission window and was pseudo-colored in red.

**Sequence analysis of DIGs and their homologs**

Protein sequence of DIGI was used as a query to search for homologous protein sequences in *A. thaliana* by the BLASTP search tool on Ensembl Plants (65). The resulting six protein sequences (Q6FK36, Q6SFMP6, Q6FGV7, Q6LK28, Q6FK56, Q6FK57) were used to query *A. thaliana*, Glycine max, *Solanum lycopersicum*, *Oryza sativa* japonica, and 21 homologous sequences. These sequences were aligned using MEGAG (66) using distance-based maximum likelihood method, and bootstrap values were generated from 1000 replications.

**Chlorophyll measurement**

Chlorophyll content was determined as previously described (67). Briefly, each sample consisting of ~50 seeds were germinated and grown on LS plates supplemented with or without ABA and DEX for eight days. The seedlings were collected and ground in liquid nitrogen. Chlorophyll were extracted by 80% acetone until pellets were almost white. Absorbance was measured at 647 and 664 nm in a DU-730 spectrophotometer (Beckman Coulter, CA, USA). Chlorophyll content was determined as

$$chl = a + b = 17.76*{A}_{647} + 7.34*{A}_{664}$$

Chlorophyll content of each transgenic line was then normalized by the corresponding seedlings grown on LS plates containing no DEX or ABA. The 95% confidence interval around the mean estimate was calculated from three biological replicates.

REFERENCES AND NOTES


A transcription factor hierarchy defines an environmental stress response network

Liang Song, Shao-shan Carol Huang, Aaron Wise, Rosa Castanon, Joseph R. Nery, Huaming Chen, Marina Watanabe, Jerushah Thomas, Ziv Bar-Joseph and Joseph R. Ecker

**Science** 354 (6312), aag1550.
DOI: 10.1126/science.aag1550

**Complex transcription factor interactions**

To respond to environmental changes, such as drought, plants must regulate numerous cellular processes. Working in the model plant *Arabidopsis*, Song et al. profiled the binding of 21 transcription factors to chromatin and mapped the complex gene regulatory networks involved in the response to the plant hormone abscisic acid. The work provides a framework for understanding and modulating plant responses to stress.

*Science*, this issue p. 598