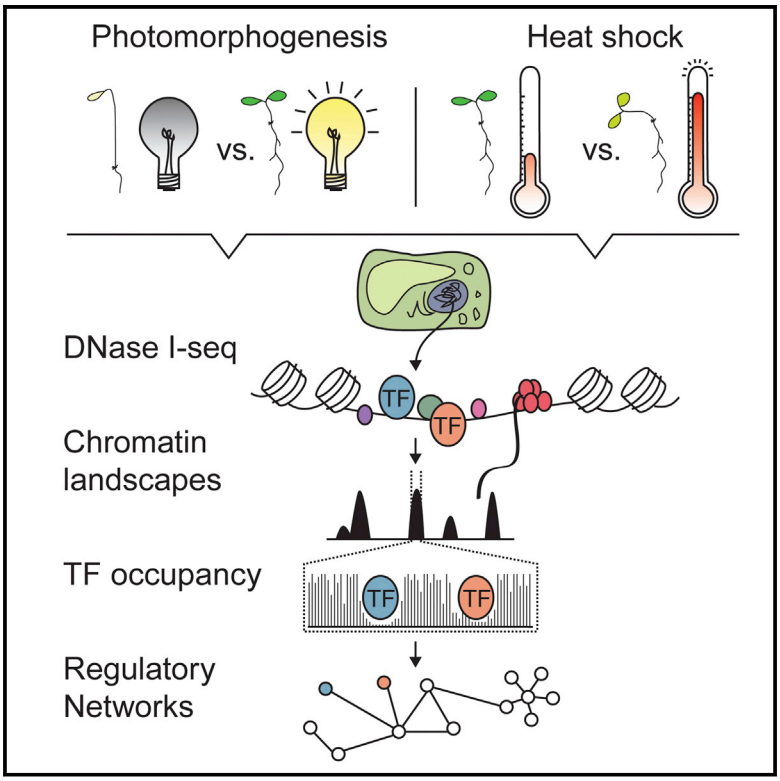
Resource

Cell Reports

Mapping and Dynamics of Regulatory DNA and Transcription Factor Networks in *A. thaliana*

Graphical Abstract



Authors

Alessandra M. Sullivan, Andrej A. Arsovski, ..., Christine Queitsch, John A. Stamatoyannopoulos

Correspondence

queitsch@uw.edu (C.Q.),
jstam@uw.edu (J.A.S.)

In Brief

Our understanding of plant gene regulation is constrained by our limited knowledge of plant *cis*-regulatory DNA and its dynamics in response to environmental cues. Sullivan et al. now establish nucleotide-resolution regulatory DNA landscapes for *A. thaliana* seedlings before and after exposure to light and heat, key environmental cues shaping plant growth and development. This study generates genome-wide, condition- and tissue-specific maps of TF occupancy, constructs condition-specific TF networks, and identifies hundreds of de novo TF motif models.

Highlights

A. thaliana regulatory DNA, TF footprints, and *cis*-regulatory lexicon are elucidated

TF binding in protein-coding exons may have shaped *A. thaliana* codon usage

A. thaliana TF network architecture is strikingly similar to human

Light- and heat-cued regulatory DNA dynamics and TF network remodeling are revealed

Accession Numbers

GSE53322

Sullivan et al., 2014, Cell Reports 8, 2015–2030
September 25, 2014 ©2014 The Authors
http://dx.doi.org/10.1016/j.celrep.2014.08.019

Cell Reports
Resource

Mapping and Dynamics of Regulatory DNA and Transcription Factor Networks in *A. thaliana*

Alessandra M. Sullivan,[1,6] Andrej A. Arsovski,[2,6] Janne Lempe,[1,6] Kerry L. Bubb,[1,6] Matthew T. Weirauch,[3] Peter J. Sabo,[1] Richard Sandstrom,[1] Robert E. Thurman,[1] Shane Neph,[1] Alex P. Reynolds,[1] Andrew B. Stergachis,[1] Benjamin Vernot,[1] Audra K. Johnson,[1] Eric Haugen,[1] Shawn T. Sullivan,[1] Agnieszka Thompson,[1] Fidencio V. Neri III,[1] Molly Weaver,[1] Morgan Diegel,[1] Sanie Mnaimneh,[4] Ally Yang,[4] Timothy R. Hughes,[4,5] Jennifer L. Nemhauser,[2] Christine Queitsch,[1,*] and John A. Stamatoyannopoulos[1,*]

[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[2]Department of Biology, University of Washington, Seattle, WA 98195, USA
[3]Center for Autoimmune Genomics and Etiology (CAGE) and Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
[4]Donnelly Centre and Department of Molecular Genetics, University of Toronto, Toronto ON M5S 3E1, Canada
[5]Canadian Institute for Advanced Research (CIFAR) Program in Genetic Networks, Toronto ON M5G 1Z8, Canada
[6]Co-first author
*Correspondence: queitsch@uw.edu (C.Q.), jstam@uw.edu (J.A.S.)
http://dx.doi.org/10.1016/j.celrep.2014.08.019

SUMMARY

Our understanding of gene regulation in plants is constrained by our limited knowledge of plant *cis*-regulatory DNA and its dynamics. We mapped DNase I hypersensitive sites (DHSs) in *A. thaliana* seedlings and used genomic footprinting to delineate ~700,000 sites of in vivo transcription factor (TF) occupancy at nucleotide resolution. We show that variation associated with 72 diverse quantitative phenotypes localizes within DHSs. TF footprints encode an extensive *cis*-regulatory lexicon subject to recent evolutionary pressures, and widespread TF binding within exons may have shaped codon usage patterns. The architecture of *A. thaliana* TF regulatory networks is strikingly similar to that of animals in spite of diverged regulatory repertoires. We analyzed regulatory landscape dynamics during heat shock and photomorphogenesis, disclosing thousands of environmentally sensitive elements and enabling mapping of key TF regulatory circuits underlying these fundamental responses. Our results provide an extensive resource for the study of *A. thaliana* gene regulation and functional biology.

INTRODUCTION

As sessile organisms, plants are shaped by their environment and respond acutely to cues such as light and temperature. Such responses result in significant alterations in gene expression; however, the *cis*-regulatory elements and transcription factor regulatory networks controlling these changes remain largely undefined.

DNase I hypersensitive sites (DHSs) (Wu et al., 1979) are the hallmark of regulatory DNA in eukaryotic genomes, and DNase I hypersensitivity mapping (Thurman et al., 2012) and genomic footprinting (Neph et al., 2012b) have been extensively employed to delineate *cis*-regulatory DNA and transcription factor (TF) occupancy at nucleotide resolution in higher organisms. Such maps have provided wide-ranging insights into genome function, evolution, and the genetic basis of common phenotypes (Maurano et al., 2012). Global mapping of transcription factor footprints provides a powerful foundation for construction of extensive regulatory networks encompassing hundreds of TFs and comparative analysis of regulatory network dynamics (Neph et al., 2012a).

Here, we apply these powerful approaches to delineate the regulatory DNA landscape of the reference plant *A. thaliana* at unprecedented resolution; to analyze the relationship between regulatory DNA and phenotype-associated variation; to define the major features of the *A. thaliana* TF lexicon and regulatory network architecture; and to map the regulatory circuitry underlying responses to temperature and light, the most important environmental cues shaping plant growth and development. All raw and processed data are available at http://plantregulome.org.

RESULTS

Distribution and Features of *A. thaliana* DHSs

To extract nuclei from *A. thaliana* tissues, we created an INTACT (Deal and Henikoff, 2010) line constitutively expressing a nuclear pore biotin tag and developed a protocol for gentle mechanical disruption of plant tissue to release nuclei, which were isolated using streptavidin affinity reagents (Supplemental Experimental Procedures). We then adapted previous DNase sequencing (DNase-seq) protocols (John et al., 2011) to create, map, and sequence DNase I fragment libraries, following which we modified DHS and footprint detection algorithms for use in the smaller *A. thaliana* genome with appropriate false discovery rate (FDR) thresholds (Supplemental Experimental Procedures).

We first performed standard-depth DNase-seq on whole-plant seedlings and confirmed that the resulting maps visualized DHSs present in more specialized subsamples such as root

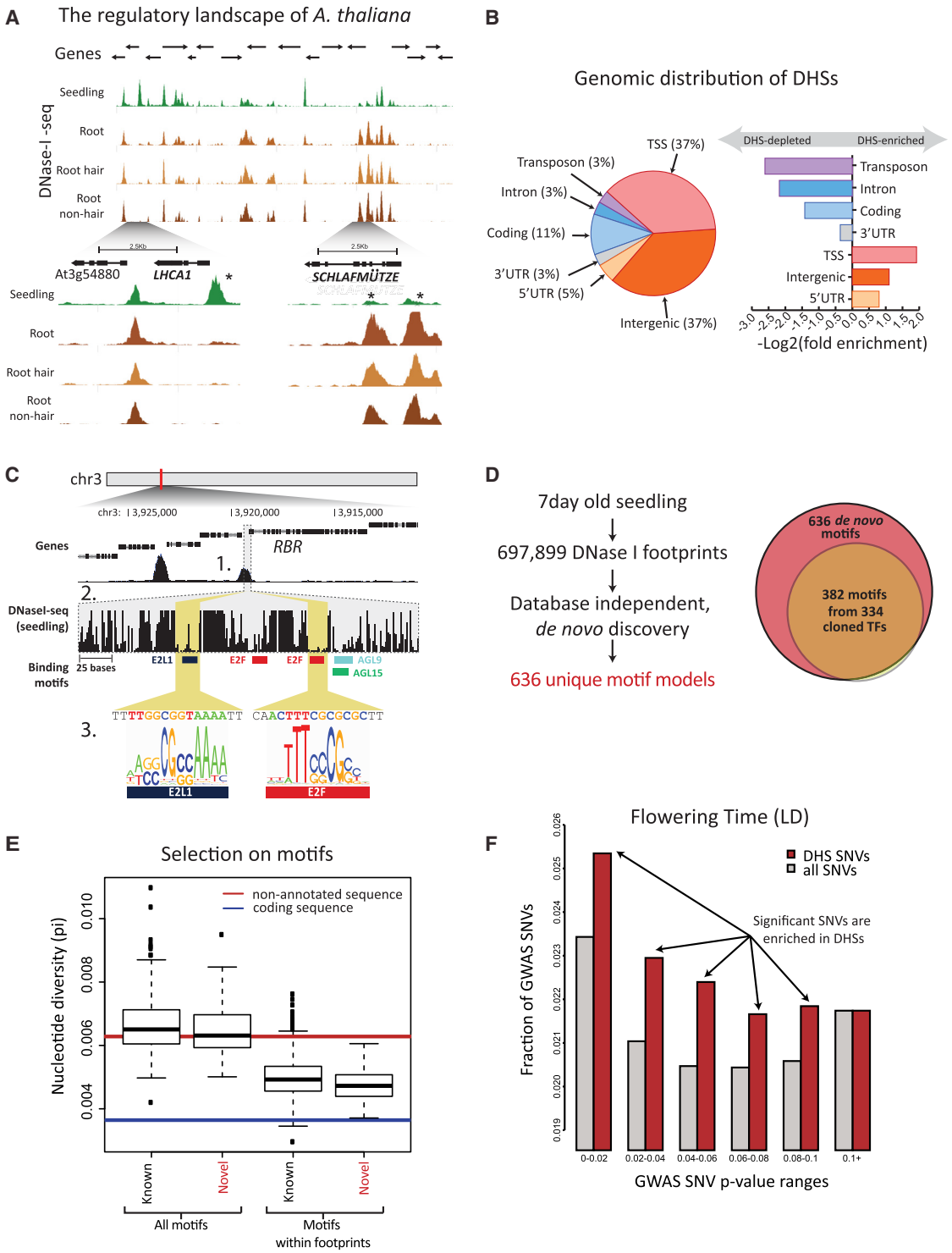


Figure 1. Nucleotide-Resolution Mapping of Regulatory DNA Enabled Discovery of TF Footprints and De Novo TF Motifs

(A) DNase I hypersensitivity (read-depth normalized density tracks) in whole seedling, seedling root tissue, and two-root epidermal cell types within a representative 100 kb region of chromosome 3. Tissue-specific DNase I hypersensitive sites (DHSs) resided near the light- and flowering-time-associated genes *LHCA1* and *SCHLAFMÜTZE* (marked by asterisk *).

(B) DHSs disproportionately resided in intergenic, TSS, and 5′ UTR elements.

(C) Representative example of footprints with TF motifs. Shown is a chromosome 3 region with tracks denoting (1) a DHS in the *RBR* promoter, (2) per-base cleavage in the *RBR* DHS with bars indicating TF motifs, and (3) footprints containing the motifs of E2L1 and E2F.

(legend continued on next page)

tissue or root epidermal cell types (e.g., root hair cells and root nonhair cells), while also revealing quantitative differences in accessibility at individual elements (Figure 1A; Table S1).

We then performed deep DNase I-seq (>260 million uniquely mapped genomic reads) on high-quality seedling samples and defined 34,288 DHSs at a stringent FDR threshold (FDR 1%), which covered 4% of the *A. thaliana* genome. The DHS distribution across the *A. thaliana* genome reflected its high gene density, with 37% of DHSs localizing within ~400 bp upstream of transcription start sites (TSSs) (Figure 1B). DHSs were relatively enriched in intergenic regions and 5′ UTRs (Figure 1B; Table S2). Although *A. thaliana* transposons and introns were generally depleted for DHSs (Figure 1B), intronic DHSs were more likely to reside within genes encoding transcriptional regulators such as the *PHYTOCHROME-INTERACTING FACTOR* 1 and 4 (*PIF*s) and *ELONGATED HYPOCOTYL 5* (*HY5*) ($p < 2.2 \times 10^{-16}$; binomial distribution test), consistent with known intron-dependent regulation of some plant TFs (Sieburth and Meyerowitz, 1997; Table S3).

A. thaliana DHSs were depleted for DNA methylation (Lister et al., 2008; $p < 1.0 \times 10^{-50}$; binomial distribution test) but contained altered ratios of cytosine methylation contexts (CG, CHG, and CHH; H indicates a non-G base), specifically, the proportion of cytosine methylation (me-C) in the asymmetric CHH context increased ($p < 2.2 \times 10^{-16}$; chi-square; Figure S1A; Table S4). All DHSs, regardless of whether they coincided with transposable elements or repeats, had similar ratios of cytosine methylation contexts. In plants, asymmetric methylation is maintained by constant de novo methylation and silences repeated and foreign DNA, including transposons (Law and Jacobsen, 2010). The increased presence of me-C in the plastic CHH context in DHSs is consistent with involvement of these regions in dynamic gene regulation.

As previously reported in other organisms (Hesselberth et al., 2009; Neph et al., 2012b), TF occupancy mapped using orthogonal approaches such as chromatin immunoprecipitation sequencing (ChIP-seq) localized within DHSs. For example, reproducible *A. thaliana* PIF3 ChIP-seq peaks (Zhang et al., 2013) were far more likely to colocalize with DHSs than nonreproducible ones (Figure S1B), and PIF3 motifs were highly enriched in reproducible ChIP-seq peaks that colocalized with DHSs (Figures S1C and S1D).

Footprinting the *A. thaliana* Genome

Dense mapping of DNase I cleavages enables genome-wide mapping of TF footprints (Hesselberth et al., 2009; Neph et al., 2012b). We defined 697,899 footprints at 1% FDR in a deeply sequenced seedling sample. Distinct footprints were readily apparent in whole seedling data and could be resolved to specific TF recognition sequences defined by classical footprinting assays (occupancy of the photomorphogenesis master regulator HY5 within the *RBCS1A* promoter; Chattopadhyay et al., 1998; Figure S2E). More-complicated relationships were also discernable. For example, footprints in the promoter of the cell cycle control gene *RETINOBLASTOMA-RELATED* (*RBR*) coincided with binding sites for the cell cycle control TFs E2F and E2L1 (Gutierrez, 2009; Figure 1C). *RBR* is the plant homolog of the human retinoblastoma gene, the protein product of which targets and inactivates E2F transcription factors (Gutierrez, 2009). The footprint data suggest a feedback loop in which *RBR* expression is regulated by its E2F targets (Vandepoele et al., 2005).

Within TF recognition sequences, per nucleotide DNA accessibility is heterogeneous and tracks the topology of the protein-DNA interface (Hesselberth et al., 2009). This feature was evident for plant-specific TFs such as ATERF-1 (Figure S1E). Two of the originally defined MADS box factors, the *A. thaliana* homeotic factor AGAMOUS and the human cell cycle regulator serum response factor share similar DNase I cleavage profiles, suggesting that DNA accessibility patterns recapitulate DNA-binding domain conservation (Figure S1F).

Expanding the *A. thaliana cis*-Regulatory Lexicon

TF footprints reflect occupancy of recognition elements by their cognate TFs. They can be systematically mined to derive the *cis*-regulatory lexicon for an organism (Neph et al., 2012b). We performed de novo motif discovery on the 697,899 well-defined (FDR 1%) seedling footprints. We identified a total of 636 distinct 8–16 bp motifs, each of which was detected in at least 1,466 footprints (median 4,799 footprints; Figure 1D; Table S5). These 636 motifs accounted for more than 89% of the seedling footprints and encompassed 80/82 of previously defined plant TF recognition sequences (Bryne et al., 2008; Matys et al., 2006).

To validate the footprint-derived motifs, we compared them with 382 motif models derived from protein-binding microarray (PBM) analysis of 334 cloned *A. thaliana* TFs (Franco-Zorrilla et al., 2014; Weirauch et al., 2014; Table S6). Of the experimentally defined motif models, 96% (366/382) were close matches to at least one of our footprint-derived motifs (Figure 1D).

To distinguish novel footprint-derived recognition sequences from known motifs, we subtracted all motif models that resembled any known TRANSFAC or JASPAR or PBM motif using liberal matching criteria (Gupta et al., 2007; Experimental Procedures), which yielded 112 novel motif models (Table S5).

To validate these 112 motifs, we analyzed recent evolutionary selection within these elements using nucleotide diversity data for 80 *A. thaliana* accessions (Cao et al., 2011). Similar to known motifs, the novel 112 footprint-derived motifs showed significantly reduced nucleotide diversity relative to neutrally evolving sequences (Figure 1E), compatible with recent evolutionary selection. These results indicate that TF footprints

(D) De novo motif discovery yielded 636 motifs. We validated these by comparing them to 382 protein-binding microarray-derived motif models (Table S6), considering only the best de novo motif match. Three hundred and sixty-six (96%) of the 382 protein-binding microarray-derived motif models matched at least one de novo motif.

(E) Nucleotide diversity was similar for de novo motifs that match known and novel motifs. Red line is nucleotide diversity of nonannotated, presumably neutrally evolving DNA. Blue line is nucleotide diversity (pi) of coding regions, which mostly evolve under purifying selection. Motifs in footprints showed substantially reduced diversity compared with all matches genome wide.

(F) Highly significant GWAS SNVs were significantly enriched in DHSs (*p value 0.00153; K-S test) for the GWAS phenotype flowering time (LD).

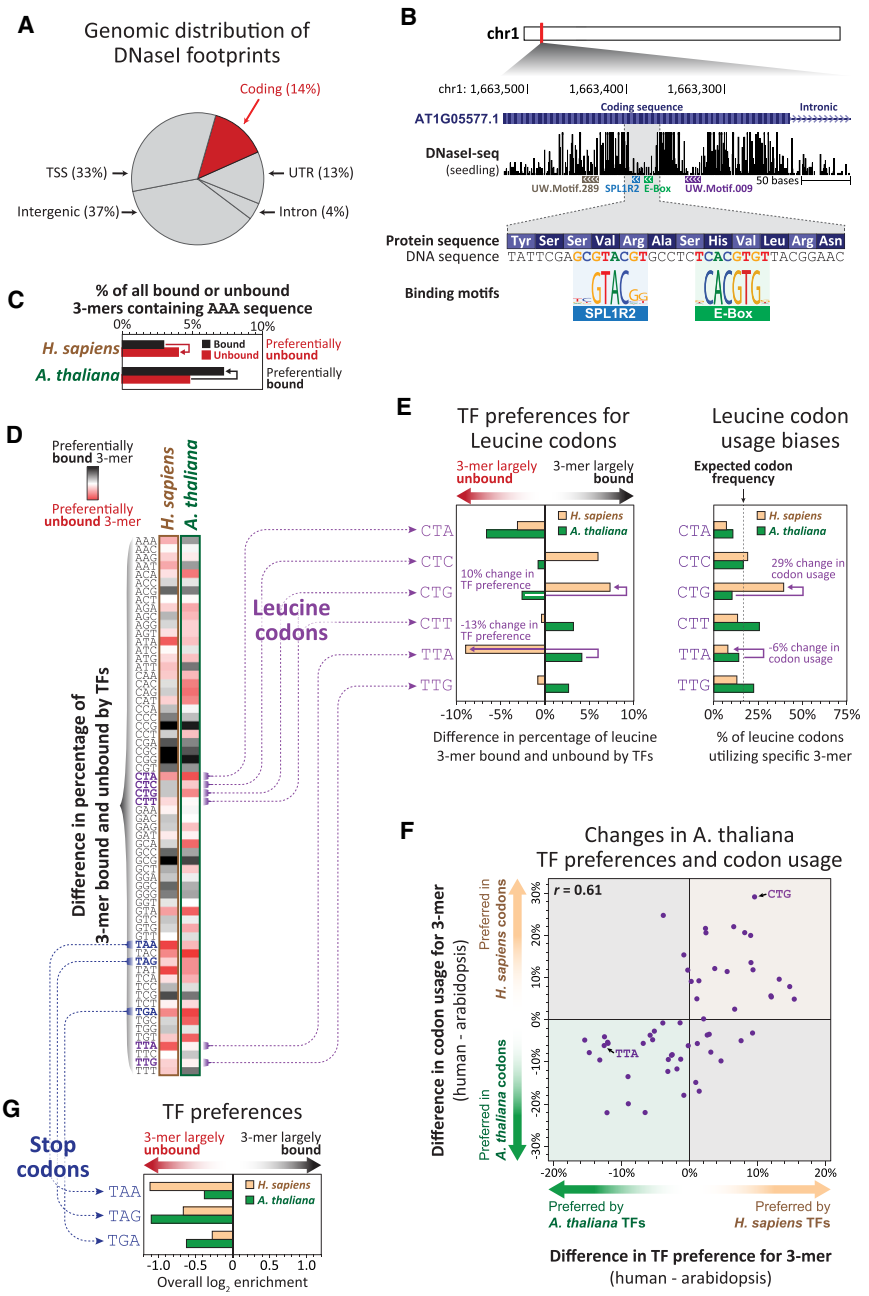


Figure 2. TF Codon Binding Preferences Correlated with Organismal Codon Bias

(A) Fourteen percent of footprints resided in coding regions.

(B) A representative example of footprints in a coding region (AT1G05577) on chromosome 1.

(C) Example of a 3-mer (AAA; encoding lysine) that was preferentially bound by TFs (i.e., resided within footprints; red) in *A. thaliana*, but not in human.

(D) TF binding preferences in *A. thaliana* and human were calculated for all codons, revealing both similarities and differences in TF binding preferences between *A. thaliana* and human.

(E) TF binding preferences for leucine codons (left) correlated with differences in *A. thaliana* and human codon usage (right).

(F) Usage differences for all codons were strongly correlated with differences in TF binding preferences.

(G) Stop codons were consistently unbound in *A. thaliana* (green) and human (light brown).

collectively define an evolving functional compartment of the *A. thaliana* genome (Figure 1E).

A. thaliana GWAS Variants Are Enriched in DHSs

In human, common disease- and phenotypic trait-associated variation mapped in genome-wide association studies (GWASs) localizes in DHSs (Maurano et al., 2012). We sought to determine whether *A. thaliana* single-nucleotide variants (SNVs) associated with diverse phenotypes (Atwell et al., 2010) showed similar properties. *A. thaliana* GWASs encompass far smaller sample sizes than human studies (<200 strains in *A. thaliana* versus thousands of humans) and are complicated by extensive population structure, leading to many false-positive associations for complex traits. Despite these limitations, we found that a significantly greater fraction of trait-associated SNVs resided within DHSs; moreover, this fraction increased with increasing SNV significance. This pattern holds for the majority (72 of 107) of GWAS phenotypes (Table S7). Flowering time is among the GWAS phenotypes exhibiting an enrichment of strongly associated SNVs in DHSs (Figure 1F; $p < 0.0015$; Kolmogorov-Smirnov [K-S] test). DHS stratification of phenotype-associated variants may thus highlight the most promising variants for functional studies.

TF Occupancy of *A. thaliana* Protein-Coding Exons May Modulate Codon Choice

In mammalian genomes, transcription factor occupancy within protein-coding exons may modulate codon choice and protein evolution (Stergachis et al., 2013). The generality of this phenomenon across kingdoms is unknown. Overall, 14% of *A. thaliana* footprints localized within protein-coding exons (Figures 2A and 2B). The specific codons that were preferentially contained in TF footprints differed substantially between *A. thaliana* and human (e.g., Lys AAA; Figures 2C–2E). Changes in TF binding preferences between *A. thaliana* and human strongly correlated with directional codon biases ($r = 0.61$; Figures 2E and 2F). For example, the leucine-encoding CTG was preferentially bound in human compared to *A. thaliana*; this codon is far more frequently utilized in human coding regions (Figure 2E). In human, stop codon trinucleotides (TAA, TAG, and TGA) are

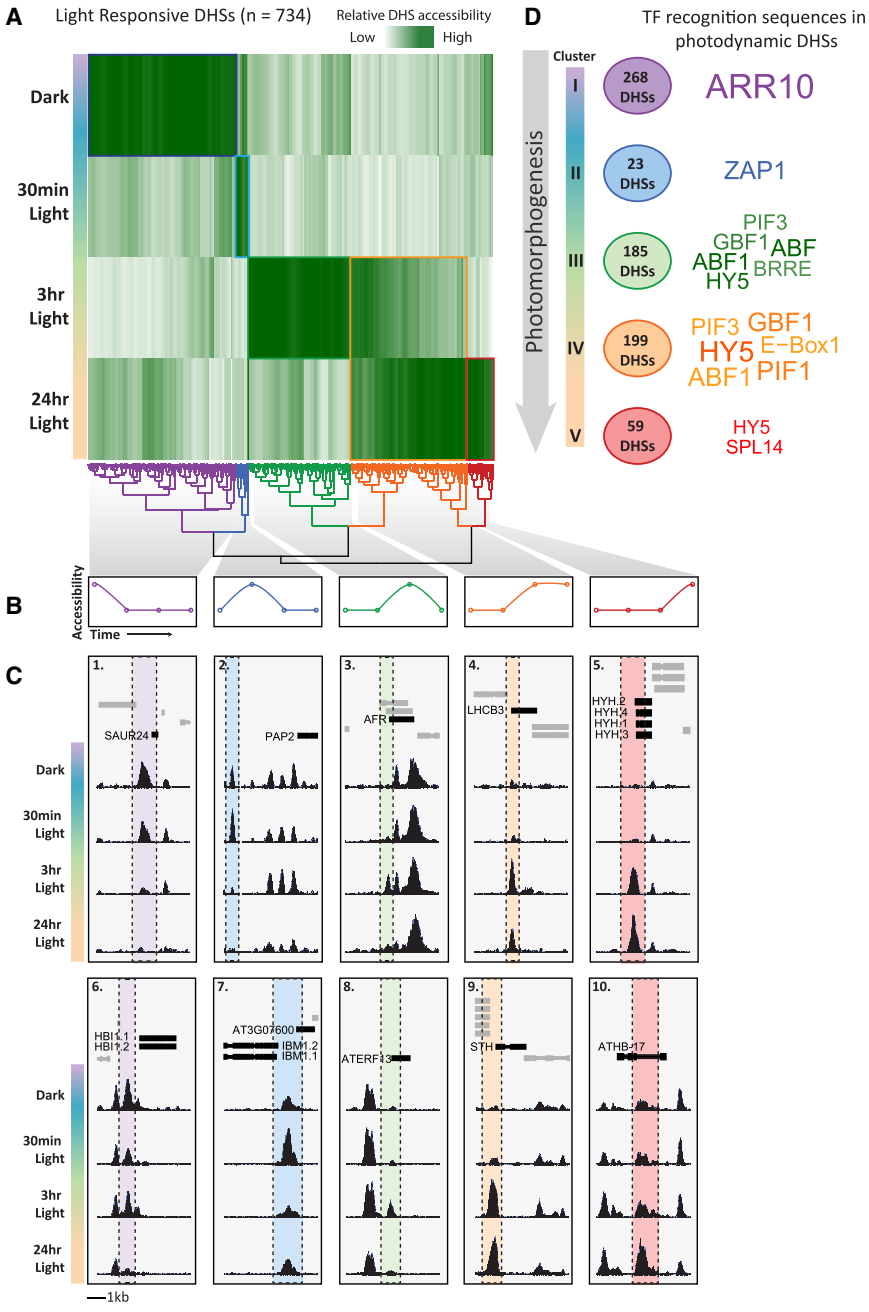


Figure 3. Dynamic Chromatin Changes during Photomorphogenesis

(A) DHSs identified for each light treatment were clustered, yielding five condition-specific DHS clusters: (I, purple) DNase I accessible in dark, (II, blue) accessible in 30 min light, (III, green) accessible in 3 hr light, (IV, orange) accessible in 3 hr and 24 hr light, and (V, red) accessible in 24 hr light.

(B) Characteristic patterns of DNase I accessibility.

(C) Representative examples of DHSs (1–10) from each condition-specific DHS cluster were located near known and novel photomorphogenesis genes. Each window is 5 kb; vertical ranges vary but are consistent for a given DHS example, highlighting fold change rather than absolute differences.

(D) TRANSFAC motif densities relative to background within each cluster are represented as a word cloud.

significantly depleted from DNase I footprints genome wide (Stergachis et al., 2013). We found an analogous situation (Figure 2G), indicating that the *A. thaliana* TF repertoire has likewise been depleted of DNA-binding domains capable of recognizing stop codons.

Regulatory DNA Landscape Dynamics during Photomorphogenesis

In seedlings, light triggers photomorphogenesis, a fundamental and irreversible reshaping of plant form and metabolism to optimize photosynthesis. Underlying this transition to photoautotrophy is a wave of transcriptional reprogramming and alteration in gross chromatin compaction (van Zanten et al., 2012), during which expression levels of nearly one third of all *A. thaliana* transcripts are altered (Ma et al., 2001).

To analyze the regulatory DNA landscape of photomorphogenesis, we exposed dark-grown seedlings to 0, 0.5, 3, or 24 hr of light (long day [LD] conditions) and performed DHS mapping and genomic footprinting at each time point. We identified 734 photodynamic DHSs across these conditions (Figures 3 and S2A–S2D; Table S8), which clustered into five distinct DHS accessibility patterns (Figure 3B). Many DHSs within the five clusters resided in proximity to genes previously implicated in the light response (Figure 3C, panels 3–5 and 9). For example, a DHS overlying the promoter of *HY5 HOMOLOG*, a key regulator of the light response (Brown and Jenkins, 2008), reached peak activity after 24 hr of light (cluster V; Figure 3C, panel 5). Numerous photodynamic DHSs were localized near genes with previously uncharacterized roles in photomorphogenesis (Figures 3C, panels 1, 2, 6–8, and 10, and S2D, panels 1–5; Table S8). For example, several members of the *SAUR* gene family, including *SAUR24* (Figure 3C, panel 1), contained dark-activated DHS in their promoters. *SAUR* genes play critical roles in cell expansion and transport of the plant hormone auxin, but their role in the dark is unknown (Spartz et al., 2012).

Gene ontology (GO) analysis of the genes proximal to photodynamic DHSs highlighted specific biological processes associated with dark- and light-activated DHSs (Table S9). Genes enriched for light stimulus and response to UV were associated

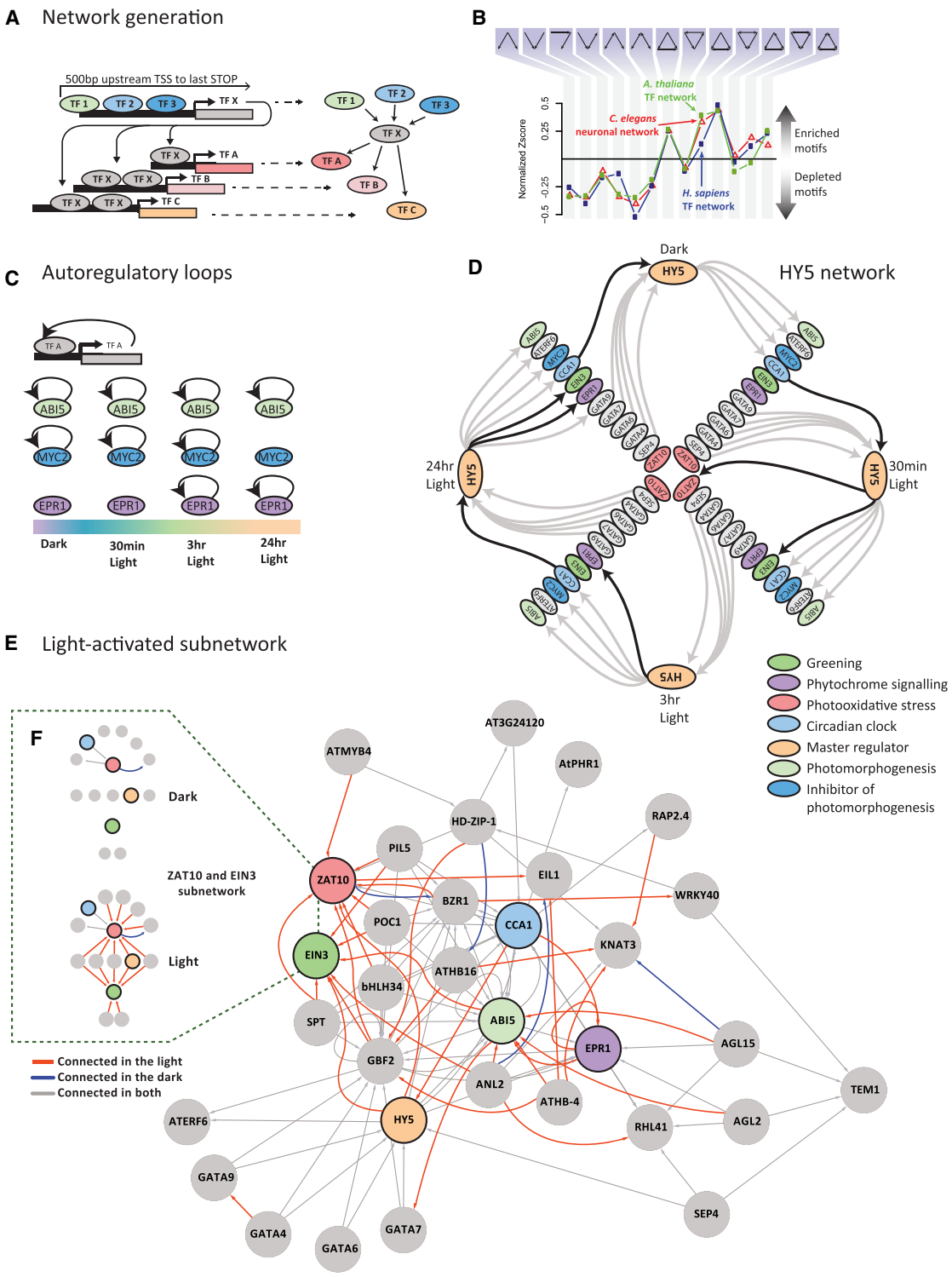


Figure 4. TF Networks Were Rewired during Photomorphogenesis

(A) To build networks, an edge was created when a footprinted motif of a source TF overlapped a target TF gene, including 500 bp upstream of the target TF's transcription start site. Target TF X in gray, TFs with regulatory input into TF X in shades of green and blue, output TFs regulated by TF X in shades of orange and red.

(B) Network motif topology in *A. thaliana* was similar to the previously described *C. elegans* neuronal network and human TF network (Milo et al., 2004; Neph et al., 2012a).

(legend continued on next page)

with 3 hr and 24 hr light-activated DHSs (cluster IV), and genes enriched for response to auxin and shade avoidance were associated with dark-activated DHSs (cluster I).

To explore the TFs mediating photodynamic DHSs, we compared the densities of recognition sequences for known TFs among DHS clusters (Figure 3D; Table S8). The light-activated DHS clusters III, IV, and V contained a high density of recognition sequences for the photomorphogenesis master regulator HY5 (Jiao et al., 2007) relative to dark-activated and 30 min light-activated DHS clusters (Figure 3D). By contrast, the light-activated DHS clusters III and IV were densely populated with PIF1 and PIF3 motifs relative to the other clusters (Figure 3D). Members of the PIF gene family control seedling growth (Leivar and Quail, 2011), and the quadruple *pif1 pif3 pif4 pif5* mutant displays a constitutive photomorphogenic phenotype (Shin et al., 2009). The dark-activated DHS cluster I contained a high density of motifs for A. THALIANA RESPONSE REGULATOR 10 (ARR10), which regulates the cytokinin response (Mason et al., 2005; Figure 3D). The cytokinin class of plant hormones is implicated in cell division, shoot initiation and growth, leaf senescence, and photomorphogenic development (Mok and Mok, 1994).

To identify which TFs distinguish photodynamic DHSs from the rest of the *cis*-regulatory landscape, we analyzed enrichment of TF recognition sequences (including novel motifs) within each cluster of photodynamic DHSs relative to all seedling DHSs, the vast majority of which were static (Figure S2F; Table S8). For the dark-specific cluster I, this analysis yielded a striking enrichment for novel motifs in addition to several ARR factors, including ARR10, and three homeobox factors (Figure S2F). Despite the fact that TFs are often lowly expressed and thus difficult to evaluate, all nine *A. thaliana* TFs with dark-enriched motifs are expressed in the dark, and six of nine, including *ARR1* and *ARR2*, show comparable or higher expression in dark relative to light conditions (Diurnal expression browser: http://diurnal.mocklerlab.org; Michael et al., 2008). The density of TF recognition sequences encoded within light-activated DHSs did not differ greatly from all seedling DHSs (Figure S2F). Indeed, when compared to the genome, recognition sequences for light-related TFs (including HY5 and PIFs) are enriched within DHSs (Figure S2G). This enrichment of light-related TF recognition motifs in all DHSs and the prevalence of static DHSs suggest that plant chromatin is poised for light.

Taken together, our results indicate that photodynamic DHSs are programmed in an exposure time-dependent fashion, which is achieved by a specific temporally coordinated set of TFs (Figure 3).

Empirical TF-to-TF Networks in *A. thaliana*

Genomic footprinting enables systematic analysis of TF occupancy within regulatory DNA of transcription factor genes, providing a direct and empirical approach for mapping cross-regulatory interactions (edges) between TF genes (nodes; Figure 4A). Systematically applying this approach to all TFs with defined recognition sequences enables the construction of TF-to-TF networks and analysis of their organization, dynamics, and architectural features (Neph et al., 2012a). This approach recapitulates validated connections, provides visualizable and interpretable information, is agnostic with respect to positive or negative interactions, and accounts for redundant recognition motifs among TFs.

To create large-scale TF regulatory networks for *A. thaliana*, we iterated the approach of Neph et al. (2012a) over 251 TFs with defined recognition sequences (from TRANSFAC, JASPAR, and PBM data). For conservative assignment of regulatory DNA to specific genes, we considered only proximal regulatory DNA (footprints occurring within 500 bp upstream of the TSS and extending over the gene body). This resulted in a network comprising 7,662 edges connecting 251 TF nodes (average of 31 edges per node). Networks are available at http://plantregulome.org.

Architecture of the *A. thaliana* TF Network

Biological networks are comprised of simple three-node network motifs that are universal and finite in number (n = 13; Milo et al., 2004). Their relative frequencies can be used to compare the topology of diverse biological networks. Analysis of network architecture is agnostic to the connection sign; i.e., any connection may be negative or positive, or both, depending on conditions. The central parameter is the connection direction ($TF_A \rightarrow TF_B$).

To analyze the architecture of the *A. thaliana* TF network, we computed the frequencies and relative enrichments of all thirteen three-node network motifs within the network. We then compared these frequencies to other biological networks, like the human TF network and the *C. elegans* neuronal network. In spite of its highly diverged *cis*- and *trans*-regulatory repertoire, this analysis revealed a highly similar topology for the *A. thaliana* TF network (Figure 4B).

Light-Induced Changes in TF Networks

Across the tested light conditions, total TF network size ranged from 1,340 regulatory edges in the dark to 1,930 edges after 3 hr of light.

Autoregulatory loops are a well-established mechanism for re-enforcing and fine-tuning gene expression patterns during

(C) Representative examples of autoregulatory loops and their dynamics in response to light.
(D) HY5 regulation with respect to selected light-related TFs (oval color denotes functional annotation) changed in response to light (black edges); however, most edges stayed constant (gray). Converging arrows represent multiple motifs found within a single footprint, which occurs when related factors can occupy the same motif within a footprint or when a single footprint spans multiple unrelated motifs. In both cases, we consider the underlying DNA to have high regulatory potential.
(E) Upon light exposure, light-related factors increased in connectivity (gray ovals; otherwise colored ovals denote functional annotation; see D). Regulatory connections observed in any of the light conditions, but not in the dark, are in red; regulatory connections observed in the dark and not in any light condition are in blue. Many regulatory connections were constitutive across all four conditions (gray).
(F) TFs related to greening (EIN3, green) and photo-oxidative stress (ZAT10, red) became highly connected in response to light and shared five regulators, including HY5 (orange).

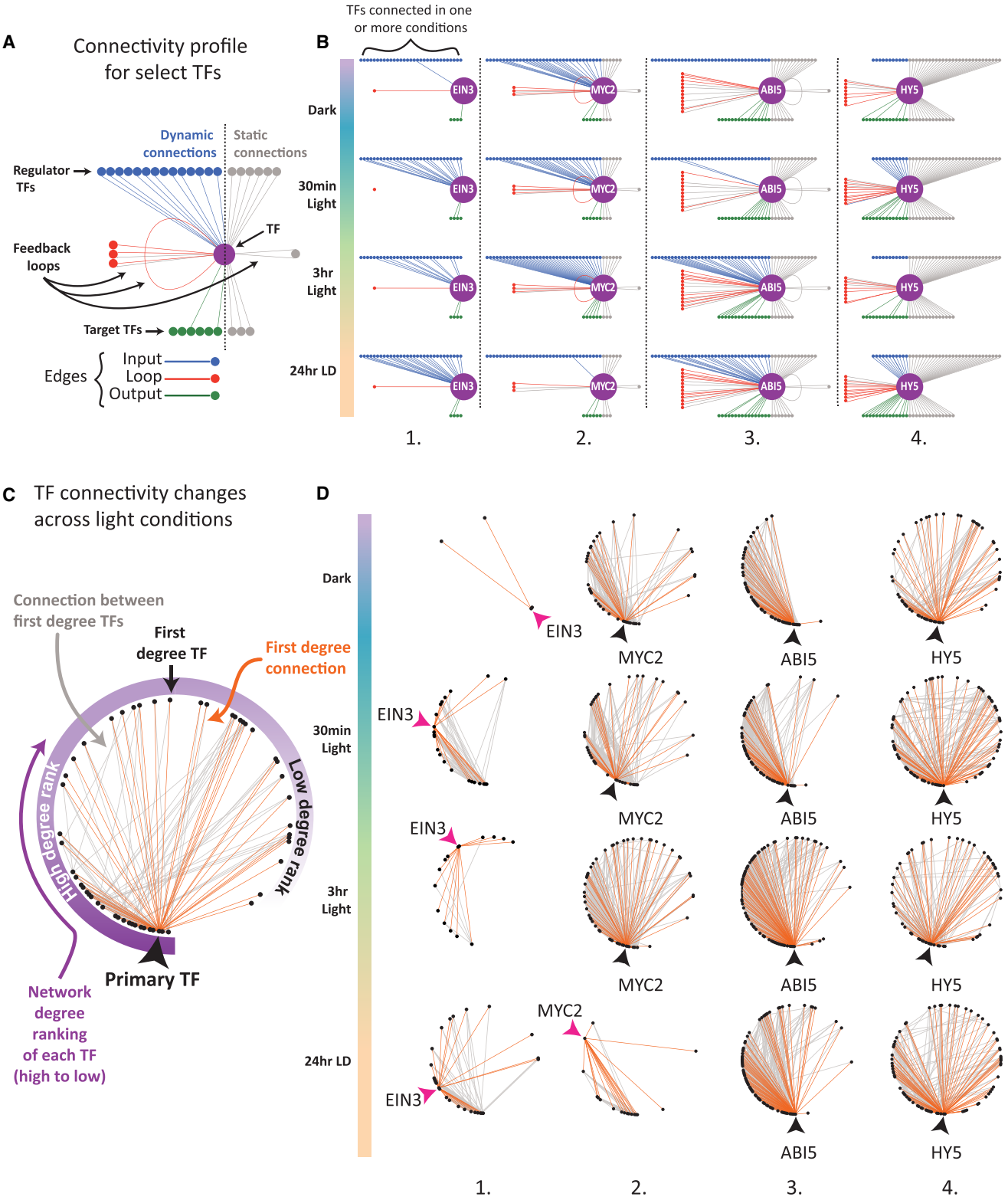


Figure 5. TF Network Rewiring followed Complex Patterns during Photomorphogenesis

(A) For representative TFs, connectivity was dissected into input edges (regulators, top), output edges (targets, bottom), bidirectional loops (condition-specific, left; common across conditions, right), and autoregulatory loops, revealing regulatory differences despite similarities in overall interaction degree (e.g., ABI5 and HY5).

(legend continued on next page)

developmental transitions (Crews and Pearson, 2009). As photomorphogenesis is one of the major developmental transitions in plant life, we reasoned that it should result in rewiring of autoregulatory loops.

We detected the appearance and disappearance of several known or posited autoregulatory loops for key photomorphogenetic factors in response to light. For example, an *EPR1* autoregulatory loop appeared with increasing exposure to light (Figure 4C). *EPR1*, which is regulated by HY5 in the light, represses expression of its endogenous copy when overexpressed (Kuno et al., 2003; Li et al., 2011). We also observed the disappearance of a *MYC2* autoregulatory feedback loop upon 24 hr of light (Figure 4C), consistent with a negative feedback loop (Dombrecht et al., 2007). MYC2, a master regulator that integrates light cues and plant-hormone-signaling pathways, acts as a negative regulator of photomorphogenesis (Yadav et al., 2005). By contrast, ABI5, a general integrator of light and abscisic acid signaling (Chen et al., 2008a), appeared to be constitutively autoregulated (Figure 4C).

We examined first-degree connections between *HY5* and several TFs implicated in photomorphogenesis (Figure 4D; Table S10). *HY5* exhibited many connections that remain stable across all conditions (Figure 4D, gray lines), in addition to dynamic connections involving known or putative photomorphogenetic factors (Figure 4D, black lines). For example, after 30 min of light, we detected *HY5* occupancy at *EIN3* and *ZAT10*, which are involved, respectively, in greening (Zhong et al., 2009) and photo-oxidative and other abiotic stresses (Rossel et al., 2007). We also identified a connection across all conditions between HY5 and *ABI5* that recapitulates genetic and biochemical evidence (Chen et al., 2008a).

We further analyzed 35 TFs with light-related GO annotations, revealing a highly interconnected subnetwork (Figure 4E), comprising 108 regulatory edges, significantly more than expected from randomly selected network TFs ($p = 0.006$). These 35 TFs were far more interconnected under light versus dark conditions (red versus blue lines, Figures 4E and 4F).

Together, these results enable the de novo identification of photomorphogenesis regulators that are dynamically rewired in response to light and previously unappreciated relationships among light-regulated TFs.

TF-Specific Photomodulated Connectivity Patterns

To gain further insight into the connectivity patterns of highly connected photomodulated TFs (EIN3, MYC2, ABI5, and HY5), we tallied all connections for each TF in the network (input and output edges and autoregulatory and bidirectional loops) and differentiated common versus dynamic edges (Figure 5A). This analysis revealed marked differences between TFs with similar baseline connectivity (Figure 5A). For example, EIN3 transitioned from a relatively unconnected dark state to a stable, highly connected light state driven chiefly by increased input edges (Figure 5B, column 1). By contrast, MYC2 showed the reverse pattern (Figure 5B, column 2). *ABI5* underwent a progressive increase in input edges with light exposure (Figure 5B, column 3), whereas the high connectivity of HY5 derived chiefly from common edges (Figure 5B, column 4).

To determine how the connectivity of a given factor was propagated to network neighbors, we compared each factor's first-degree regulatory relationships (Figure 5C, orange edges) and, in turn, the first-degree regulatory relationships shared among its first-degree neighbors (Figure 5C, gray edges), and then sorted all TFs by their overall degree of connectedness in the entire network (Figure 5C). Qualitative differences again emerged between highly connected TFs. For example, following 30 min of light, EIN3 and its first-degree neighbors became more interconnected (Figure 5D, column 1). By contrast, MYC2 showed initially increased connectivity with first-degree neighbors, yet almost all these interactions were lost in response to 24 hr light (Figure 5D, column 2). Our observations highlight that detailed network analysis can reveal striking differences among major TFs with regard to their connectivity patterns across conditions and network neighborhood.

Impact of Heat Shock on the Regulatory DNA Landscape

Exposure to heat triggers a conserved response involving the rapid upregulation of heat shock proteins (HSPs), accompanied by downregulation of many normally active genes and the inhibition of most protein translation (Lindquist, 1986). Although heat shock has been studied for decades in other organisms, little is known about the impact of heat shock on plant chromatin and transcriptional regulatory pathways outside of studies of the heat shock transcription factors (HSFs) (Scharf et al., 2012) and the nucleosome variant H2A.Z in the response to ambient temperature change (Kumar and Wigge, 2010).

To define the effects of heat on the chromatin landscape and TF regulatory network, we mapped DHSs and TF footprints in control and heat-treated 7-day-old seedlings (Experimental Procedures). We focused on the most extreme (top and bottom 2.5%) heat-activated or heat-repressed DHSs (Figures 6A, S3A, and S3B; Table S11). This approach identified equal numbers of strongly heat-activated and heat-repressed DHSs ($n = 990$; 1,980 total), but fold differences varied. The genomic distributions of heat-activated versus -repressed DHSs also differed markedly, with the former localizing in distal intergenic regions and the latter localizing primarily in gene-proximal regions (TSS, 5′ UTR, 3′ UTR, intron, and coding; Figure 6B; Table S11).

A Subset of Genes with Extreme Heat-Induced DNase I Accessibility

A small fraction (14.6%; $n = 145$) of heat-activated DHSs displayed extreme accessibility (Figure 6A) and were concentrated

(B) Connectivity changed across light conditions for key TFs: EIN3, MYC2, ABI5, and HY5.
(C) First-degree neighborhoods of the same TFs visualize how a TF's connectivity percolated through the larger network. In each circular network, TFs were sorted by number of regulatory connections, such that connectivity increases clockwise around the circle, starting from the bottom. For each TF, only edges between TF and first-degree neighbors (orange) and the edges between first-degree neighbors (gray) are shown. Factors with large condition-specific changes in connectivity (EIN3 and MYC2 at 24 hr LD) are demarcated by pink arrows.
(D) Neighborhood connectivity changed across light conditions for key TFs: EIN3, MYC2, ABI5, and HY5.

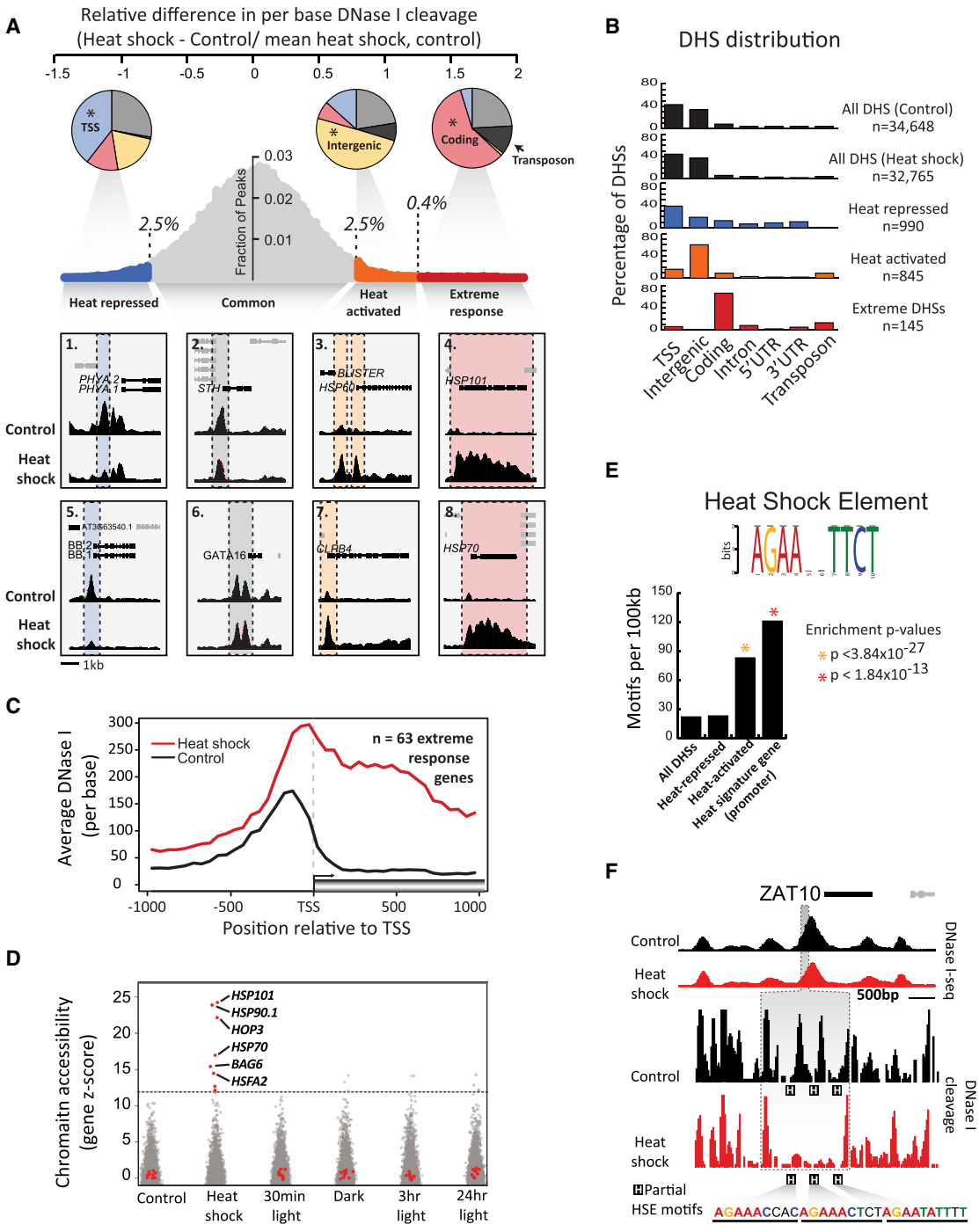


Figure 6. Dynamic Chromatin Changes in Response to Heat Shock

(A) Relative difference in DHSs between control and heat shock (Supplemental Experimental Procedures). The most heat-shock-responsive DHSs were designated as heat-activated (top 2.5%) and heat-repressed (bottom 2.5%) DHSs. The heat-activated DHSs (orange) showed a long tail (red). The most extreme heat-activated DHSs (red) encompassed extremely accessible heat-induced genes. Pie charts reflect the genomic distributions of DHS type (e.g., heat activated); asterisks (*) denote the genomic feature, in which the greatest proportion of DHSs resided relative to other genomic features.

(B) Genomic DHS distributions were similar for control and heat shock (black); however, among heat-repressed DHSs (blue), heat-activated DHSs (orange), and DHSs in promoters of extremely accessible, heat-induced genes (red) distributions varied markedly.

(C) The promoter regions of extremely accessible genes were poised for activation. Aggregated DNase I accessibility across these genes and their upstream regions is shown for control (black) and after heat shock (red). The left plot side (left of gray dotted line) shows average per-base DNase I accessibility 1,000 bp upstream of the TSS of extremely accessible, heat-induced genes. The right side (right of gray dotted line, TSS) shows normalized DNase I accessibility over the first 1,000 bp of coding regions; the average length of extremely accessible genes was 1,651 bp.

(legend continued on next page)

within the bodies of 63 genes encoding canonical heat shock proteins, their cochaperones, and several heat-stress-related TFs, in addition to novel heat-shock-responsive genes (Figures 6A–6D; Table S11). Above a *Z* score of 12 (dotted line), all extremely accessible genes were unique to heat shock (Figure 6D, red dots).

The most highly accessible gene was *HSP101*, which is crucial for acquired thermotolerance (Queitsch et al., 2000), followed by the genes encoding the heat-inducible HSP90.1 chaperone and HOP3, an important cochaperone of HSP90 (Krishna and Gloor, 2001; Figures 6A and 6D). This extreme DNase I accessibility was unique to specific members of gene families, presumably reflecting their functional specialization in the acute heat shock response.

Genes with extreme accessibility displayed "poised" promoters (Keene et al., 1981) in control conditions (Figure 6C) and tended to be highly expressed upon heat shock (Figure S3C). They were even more significantly enriched for the GO term "response to heat" (p value 4.11×10^{-71}) than the genes associated with heat-activated DHSs (p value 6.45×10^{-13}; Figure S3D). In contrast, heat-repressed DHSs were enriched near growth, transport, and metabolic genes (Figure S3D), consistent with downregulation of these energy-requiring cellular functions.

Taken together, our results identify genes displaying a unique chromatin signature with poised promoters in control conditions and extreme accessibility in response to heat shock. Given the outsize importance of heat tolerance in today's agriculture, these genes may be of direct relevance for genetic engineering.

TF Drivers of Heat-Activated DHSs

To identify TF drivers of the regulatory changes accompanying heat shock, we analyzed the TF recognition site repertoire of heat-responsive DHSs. The heat shock element (HSE) (AGAAnnTTCT) was highly enriched in heat-activated DHSs and the promoters of extreme-accessibility genes (Figures 6E and S3E; Table S11). We also detected complex, heat-associated footprinting patterns at partial HSEs. For example, the promoter of *ZAT10*, a factor involved in response to abiotic stresses that is activated by HSFA2 in a heat-shock-specific manner (Schramm et al., 2006), contains three adjacent partial HSEs, all of which showed footprints in control conditions (Figure 6F). In response to heat, these footprints merged into a single larger footprint, suggesting either that the resident HSF (presumably HSF2A) greatly increases occupancy time or that it partners with additional regulatory factors.

Unexpectedly, MADS box motifs, such as the recognition sequence for AGL9 (SEP3), were also enriched in heat-activated DHSs (Figure S3E; Table S11). Both the HSE and MADS box motifs (CArG-box) were selective for heat-activated elements, present among the top 25 overrepresented footprint-derived motifs in heat-activated DHSs but absent from the top 25 footprint-derived overrepresented motifs in heat-repressed DHSs. Similar to static DHSs during photomorphogenesis, static DHSs during heat shock did not contain vastly different TF recognition sequence compositions from all seedling DHSs; the few significant differences were of small magnitude (Figure S3C). Light-related TF motifs (HY5 and PIFs), which were generally pervasive in *cis*-regulatory DNA (Figure S2G), were among the motifs significantly enriched in static DHSs and depleted from heat-dynamic DHSs. Collectively, our findings reaffirm the importance of HSFs in driving the heat shock response and suggest a prominent role for MADS-box factors in remodeling the chromatin and regulatory landscape in response to heat shock.

Rewiring and Involution of the TF Network in Response to Heat Shock

Heat shock resulted in substantial rewiring of the TF network, with a net loss (9%) of network edges (Figure 7A). Loss of network edges was widely distributed across TF nodes rather than being confined to a few highly connected TFs (Figure 7A). These results indicate a central role for the TF network in mediating the repressive component of the heat shock response. RNA polymerase II is known to depart from actively transcribed genes in response to heat shock, yet the mechanisms driving this dynamic are unknown (Teves and Henikoff, 2011). Our findings suggest that the loss of TF occupancy at target promoters may contribute to Pol II departure and global downregulation of transcription.

HSF-Centric Subnetworks

A. thaliana encodes 21 HSFs, which fall into different classes based on domain structure. There is no single master regulator of heat-shock-responsive genes; rather, three major HSFs (HSFA1a, HSFA1b, and HSFA2) together regulate the early heat shock response (Nover et al., 2001). To gain insight into HSF regulation in response to heat shock, we analyzed HSF-centric subnetworks comprising all edges connecting the 21 *A. thaliana* HSFs in control versus heat-treated conditions. We detected only subtle changes in *HSFA1A* and *HSFA1B* regulation upon heat shock, consistent with their constitutive expression (Figure 7B). By contrast, the heat-inducible HSFs such as *HSFA2*, *HSFA7A*, *HSFA7B*, *HSFB1*, and *HSFB2B* (Kilian et al., 2007) are increasingly regulated upon heat shock (Figure 7B).

The rewiring of TF networks in response to heat shock also offered information about less well-understood HSFs, some of which function in development rather than in the canonical heat shock response (Kotak et al., 2007). For example, HSFA9 regulates heat shock gene expression during seed development, yet is not induced during heat shock; *HSFA9* showed no alteration in network connectivity. HSFA8 is induced only during

(D) Extremely accessible heat-induced genes (red dots) were unique to heat shock. Labeled genes (*HSP101*, *HSP90.1*, *HOP3*, *HSP70*, *BAG6*, and *HSFA2*) were more accessible in heat shock than similarly sized genes in any other condition examined.
(E) HSE was the most highly enriched motif within heat-shock-activated DHSs and within DHSs marking extremely accessible gene promoters relative to all DHSs (p values from Bonferroni-corrected hypergeometric tests performed on motif counts).
(F) Differential footprinting in the promoter of ZAT10, which is involved in tolerance to abiotic stresses (Mittler et al., 2006). The differentially footprinted region coincided with three partial HSE motifs.

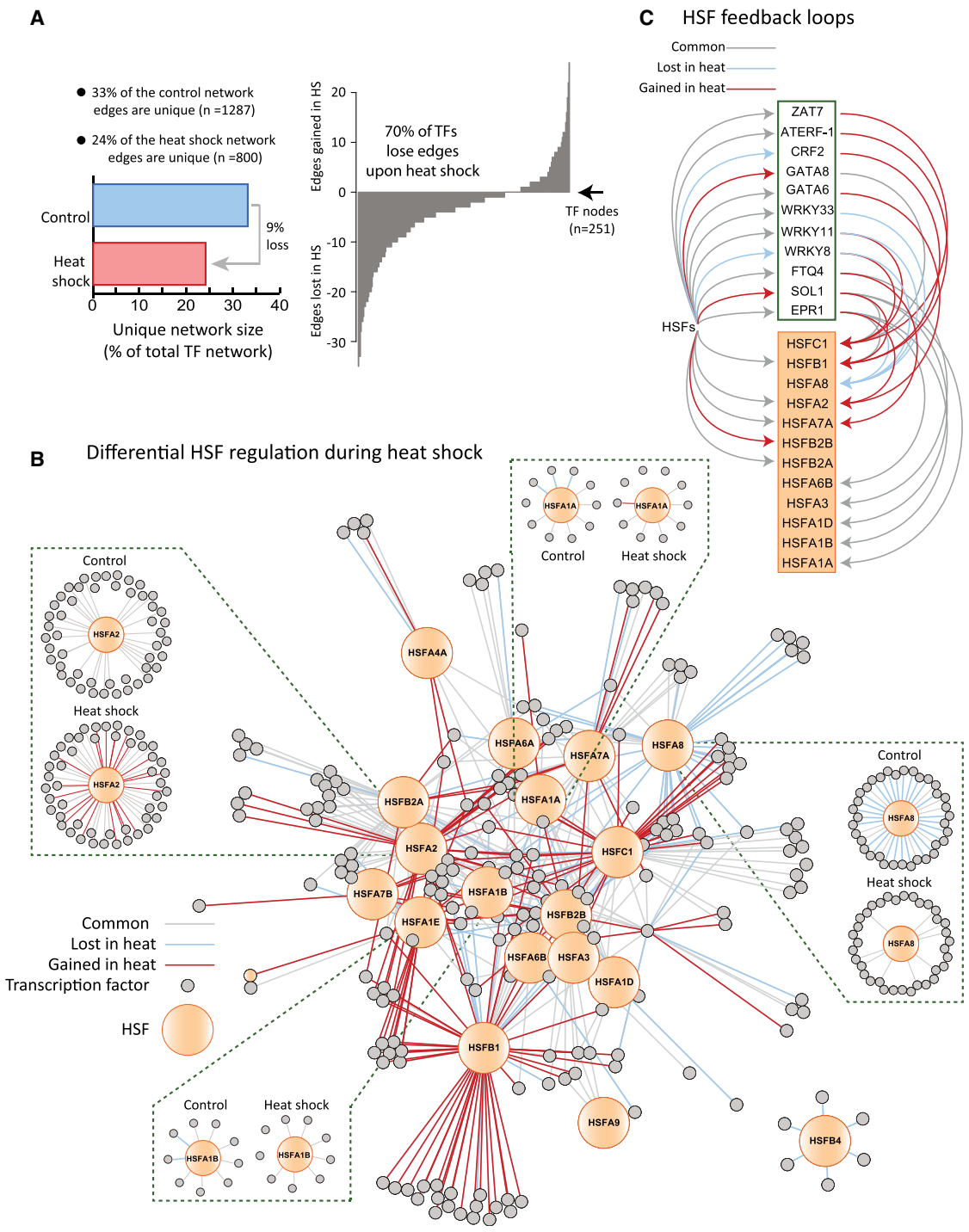


Figure 7. Network Connectivity Was Lost and HSF Networks Were Rewired in Response to Heat Shock

(A) Left: the number of unique edges in the TF network decreased in heat shock (red) compared to control (blue). Network edges were defined as in Figure 4A. Right: TFs tended to lose network edges upon heat shock.

(B) HSF family members (orange circles) were differentially regulated (input edges only). HSFA2, for example, gained many regulators in response to heat shock (see detail left, heat shock-specific edges, red). In contrast, HSFA8 lost many regulators in response to heat shock (see detail right, control-specific edges, gray). This arrangement revealed that HSFB4 was regulated by a small neighborhood of TFs that did not regulate any other HSFs.

(C) HSF feedback loops were generated, using the generic HSE motif for all HSF family members. The indicated TFs (green box) were regulated by one or more HSF and in turn regulated at least one specific HSF.

heat shock recovery; consistent with this role, *HSFA8* showed numerous regulatory edges in the control state that were lost in heat shock (Figure 7B, right).

Finally, we examined HSF feedback loops representing regulation of HSF genes by other HSFs or HSF-regulated TFs (Figure 7C). We observed the formation of several novel feedback loops upon heat shock, including a loop between the stress responsive activator ATERF-1 and the heat-activated *HSFB1* (Fujimoto et al., 2000; Figure 7C). HSFs thus form a densely connected subnetwork linked by shared first-order connections with non-HSF TFs.

DISCUSSION

We used genomic footprinting and transcription factor network analysis to map the *cis*- and *trans*-regulatory landscapes of *A. thaliana* whole seedlings and landscape dynamics in response to light and heat, the two most important environmental cues modulating plant growth and development.

First, the availability of a map of *A. thaliana* DHSs enables integration with other DNA-sequence-based annotations such as DNA methylation or genetic variation. For example, SNVs most strongly associated with diverse *A. thaliana* quantitative traits were concentrated in the ~4% of the genome marked by DHSs. Second, using genomic footprinting, we defined a core regulatory compartment in the *A. thaliana* genome comprising nearly 700,000 short sequence elements occupied in vivo by TFs. From these elements, we derived a *cis*-regulatory lexicon for *A. thaliana*, including many novel motifs that show evidence of recent purifying selection. We also find evidence that, similar to humans, TF binding shapes codon bias in *A. thaliana*. Finally, we leveraged genomic-footprinting data to construct large-scale TF-to-TF regulatory networks.

In spite of the vast divergence between plants and humans, the architecture of *A. thaliana* transcription factor network was strikingly similar to that of the human TF network and other complex information-processing systems. This conservation of non-rate-limited information processing between plants and animals is striking because plant development is so exquisitely sensitive to environmental cues. In contrast to the multicellular *A. thaliana*, other sessile organisms with acute environmental responses (the unicellular eukaryote yeast and bacteria) show rate-limited, sensory networks for quick responses to transient signals. As the ancestors of plants and animals were unicellular, our study suggests convergent multicellularity rather than environmental responsiveness as a major driver of optimal network topology.

At the *cis*-regulatory level, photomorphogenesis is characterized by a progression of distinct regulatory DNA compartments that respond to specific light exposures, each of which can be linked to the sequential actions of discrete TFs. We speculate that the prevalence of light-related TF recognition sequences in all seedling DHSs and the prevalence of static DHSs during photomorphogenesis reflect that plant chromatin is poised for the light response. At the TF network level, these changes are accompanied by substantial rewiring between groups of light- and dark-responsive TFs. This rewiring is particularly well demonstrated for autoregulatory feedback loops.

At the *cis*-regulatory level, the hallmark of heat shock is the sharp partitioning of dynamic DHSs into gene-proximal (repressed) and gene-distal (activated) compartments. We also identified heat-responsive genes that develop extreme DNase I accessibility over their promoters and gene bodies upon heat shock. A majority of the DHS landscape remained surprisingly static given the globally repressive nature of heat shock. Because chromatin remodeling is energy intensive, persistence of DHSs may facilitate rapid recovery from heat shock. At the TF network level, the large decrease in network edges suggests a role for departing TFs in mediating transcriptional repression during heat shock.

DHS and footprint data can be applied to improve interpretation of GWASs and quantitative trait locus studies by pinpointing potentially functional noncoding variants. The data can also guide the selection of mutant lines from insertion collections or the selection of DNA elements that can be targeted to perturb specific pathways. Finally, our results constitute a reference against which other *A. thaliana* accessions may be compared, as much of the vast phenotypic variation among diverse *A. thaliana* accessions is thought to arise from noncoding regulatory regions (Gan et al., 2011).

EXPERIMENTAL PROCEDURES

Plant Materials

The *UBQ10* INTACT line, in which the *UBQ10* promoter is fused with nuclear-targeting fusion protein, is available from Arabidopsis Biological Resource Center (CS68649).

Treatments

All samples were prepped at the same time of day. Seven-day-old seedlings were used for root samples.

Light Treatments

Seven-day-old dark-grown seedlings were exposed to light for 0 min, 30 min, 3 hr, and 24 hr, respectively.

Heat Treatments

Seven-day-old seedlings were heat shocked at 45°C for 30 min; control plants remained in LD conditions.

Sample Preparation

Nuclei were purified and treated with 45u DNase I for 3 min at 25°C. Size fractionation and sequencing of double-cut DNA fragments were done as described (Hesselberth et al., 2009; Neph et al., 2012b). RNA was extracted from 100–200 mg tissue, treated with DNase I, and subjected to ribosomal subtraction before library prep and sequencing. RNA expression differences were determined with Cufflinks and Cuffdiff 2.0.2 (Trapnell et al., 2012). Short read archives are in Gene Expression Omnibus (GEO) accession GSE53322. For detailed protocols, see http://plantregulome.org/protocols.

Mapping DNase I Hypersensitivity

Uniquely mapping sequencing reads (36 bp) were mapped to TAIR9. The 5′ ends of reads were used to calculate per-base DNase I cleavage. DNase-I-sensitive regions (hot spots) and DHSs (150 bp peaks) were identified as in John et al. (2011) with minor modifications. Read depth was normalized by subsampling reads. Footprints were computed as described previously (Neph et al., 2012b).

General Features of the Chromatin Landscape

Binomial distribution tests determined the probability of DHS and/or footprint overlaps with genomic features, including introns and methylated cytosines. Chi-square tests determined whether ratios of cytosine methylation contexts (CG:CHG:CHH) within DHSs deviated significantly from ratios in regions

outside of DHSs. Dark-grown seedling data were used to determine DHS and footprint overlaps with previously published PIF3 ChIP-seq data (Zhang et al., 2013).

De Novo Motif Discovery

Six hundred thirty-six de novo motifs were discovered by clustering sequences found within footprints from a deeply sequenced 7-day-old seedling sample. De novo motifs were validated in two ways: first by comparing them to 382 protein binding microarray-derived motif models (see Supplemental Experimental Procedures) and second by comparing estimates of π for neutrally evolving DNA, DNA under purifying selection, and known and novel motifs within and outside of 1% FDR footprints.

GWAS

For each GWAS phenotype, we used a nonparametric, one-sample Kolmogorov-Smirnov test to determine if low p value SNVs were more likely to occur in DHSs.

TF Trinucleotide Preferences and Codon Usage Bias

Trinucleotide frequencies within footprinted and nonfootprinted regions were tabulated for coding and noncoding portions of the human and *A. thaliana* genomes. Codon usage was determined from consensus coding sequence gene annotations (Pruitt et al., 2009) in human and coding sequences listed in the TAIR10_GFF3_genes.gff file in *A. thaliana*.

Dynamic DHSs

In the light series, we defined dynamic DHSs as the top 2% most variable DHSs across conditions. In heat shock, we defined dynamic DHSs as the 5% of DHSs with the greatest relative difference between conditions.

Motif Enrichment

Hypergeometric tests were used to test if motifs' frequencies differed in DHS subsets.

Networks

Methods are as in Neph et al. (2012a), except the region scanned for TF motifs within footprints included 500 bp upstream of the TSS and the entire gene model. Potential TF-binding sites were determined using Find Individual Motif Occurrences (FIMO) (Bailey et al., 2009), version 4.6.1, with a maximum p value threshold of 10^{-4} and defaults for other parameters.

Light-Activated Subnetworks

We simulated the random selection of 35 light-related TFs' edges from the network without replacement 1,000 times to test the significance of finding 108 regulatory edges among 35 TFs.

First-Degree TF Neighborhoods

Using Cytoscape (Shannon et al., 2003), all first-degree nodes and adjacent edges were selected and all other nodes were removed.

DNase-I-Accessibility Gene Outliers

DNase-I-accessibility gene outliers were identified by calculating *Z* scores from DNase I cleavages overlapping TAIR10 genes.

HSF Regulation during Heat Shock

HSF-centric regulatory networks were constructed by scanning HSF gene-regulatory regions for TF motifs within footprints. The generic HSF motif (Megraw and Hatzigeorgiou, 2010) was used to represent any of the 21 possible HSFs when calculating HSF feedback edges. For details, see Supplemental Experimental Procedures.

ACCESSION NUMBERS

The GEO accession number for short read data reported in this paper is GSE53322.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and eleven tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.08.019.

AUTHOR CONTRIBUTIONS

J.L., J.L.N., C.Q., and J.A.S. conceived the project. A.M.S., A.A.A., and others executed DNase I-seq experiments. A.M.S., K.L.B., and others performed data analysis. A.M.S, C.Q., and J.A.S. wrote the manuscript. All authors read, commented on, and approved the manuscript.

ACKNOWLEDGMENTS

This work was supported by grants from the National Science Foundation (MCB1243627; to J.A.S., C.Q., and J.L.N.), Graduate Research Fellowship (DGE-0718124; to A.M.S.), and EMBO long-term and HFSP long-term fellowships (to J.L.). We thank Roger Deal and Steven Henikoff for sharing INTACT lines and experimental expertise, members of the J.A.S. and C.Q. labs for useful discussions, and Stanley Fields for helpful comments on the manuscript.

Received: December 4, 2013
Revised: May 20, 2014
Accepted: August 7, 2014
Published: September 11, 2014

REFERENCES

Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., et al. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature *465*, 627–631.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208.

Brown, B.A., and Jenkins, G.I. (2008). UV-B signaling pathways with different fluence-rate response profiles are distinguished in mature Arabidopsis leaf tissue by requirement for UVR8, HY5, and HYH. Plant Physiol. *146*, 576–588.

Bryne, J.C., Valen, E., Tang, M.H.E., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., and Sandelin, A. (2008). JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res. *36*, D102–D106.

Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., et al. (2011). Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat. Genet. *43*, 956–963.

Chattopadhyay, S., Ang, L.H., Puente, P., Deng, X.W., and Wei, N. (1998). Arabidopsis bZIP protein HY5 directly interacts with light-responsive promoters in mediating light control of gene expression. Plant Cell *10*, 673–683.

Chen, H., Zhang, J., Neff, M.M., Hong, S.W., Zhang, H., Deng, X.W., and Xiong, L. (2008a). Integration of light and abscisic acid signaling during seed germination and early seedling development. Proc. Natl. Acad. Sci. USA *105*, 4495–4500.

Crews, S.T., and Pearson, J.C. (2009). Transcriptional autoregulation in development. Curr. Biol. *19*, R241–R246.

Deal, R.B., and Henikoff, S. (2010). A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev. Cell *18*, 1030–1040.

Dombrecht, B., Xue, G.P., Sprague, S.J., Kirkegaard, J.A., Ross, J.J., Reid, J.B., Fitt, G.P., Sewelam, N., Schenk, P.M., Manners, J.M., and Kazan, K. (2007). MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. Plant Cell *19*, 2225–2245.

Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors

and their potential to define target genes. Proc. Natl. Acad. Sci. USA *111*, 2367–2372.

Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H., and Ohme-Takagi, M. (2000). Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. Plant Cell *12*, 393–404.

Gan, X., Stegle, O., Behr, J., Steffen, J.G., Drewe, P., Hildebrand, K.L., Lyngsoe, R., Schultheiss, S.J., Osborne, E.J., Sreedharan, V.T., et al. (2011). Multiple reference genomes and transcriptomes for Arabidopsis thaliana. Nature *477*, 419–423.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol. *8*, R24.

Gutierrez, C. (2009). The Arabidopsis cell division cycle. Arabidopsis Book *7*, e0120.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., et al. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat. Methods *6*, 283–289.

Jiao, Y., Lau, O.S., and Deng, X.W. (2007). Light-regulated transcriptional networks in higher plants. Nat. Rev. Genet. *8*, 217–230.

John, S., Sabo, P.J., Thurman, R.E., Sung, M.H., Biddie, S.C., Johnson, T.A., Hager, G.L., and Stamatoyannopoulos, J.A. (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat. Genet. *43*, 264–268.

Keene, M.A., Corces, V., Lowenhaupt, K., and Elgin, S.C. (1981). DNase I hypersensitive sites in Drosophila chromatin occur at the 5′ ends of regions of transcription. Proc. Natl. Acad. Sci. USA *78*, 143–146.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J., and Harter, K. (2007). The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. Plant J. *50*, 347–363.

Kotak, S., Vierling, E., Bäumlein, H., and von Koskull-Döring, P. (2007). A novel transcriptional cascade regulating expression of heat stress proteins during seed development of Arabidopsis. Plant Cell *19*, 182–195.

Krishna, P., and Gloor, G. (2001). The Hsp90 family of proteins in Arabidopsis thaliana. Cell Stress Chaperones *6*, 238–246.

Kumar, S.V., and Wigge, P.A. (2010). H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. Cell *140*, 136–147.

Kuno, N., Møller, S.G., Shinomura, T., Xu, X., Chua, N.H., and Furuya, M. (2003). The novel MYB protein EARLY-PHYTOCHROME-RESPONSIVE1 is a component of a slave circadian oscillator in Arabidopsis. Plant Cell *15*, 2476–2488.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220.

Leivar, P., and Quail, P.H. (2011). PIFs: pivotal components in a cellular signaling hub. Trends Plant Sci. *16*, 19–28.

Li, G., Siddiqui, H., Teng, Y., Lin, R., Wan, X.Y., Li, J., Lau, O.S., Ouyang, X., Dai, M., Wan, J., et al. (2011). Coordinated transcriptional regulation underlying the circadian clock in Arabidopsis. Nat. Cell Biol. *13*, 616–622.

Lindquist, S. (1986). The heat-shock response. Annu. Rev. Biochem. *55*, 1151–1191.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell *133*, 523–536.

Ma, L., Li, J., Qu, L., Hager, J., Chen, Z., Zhao, H., and Deng, X.W. (2001). Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways. Plant Cell *13*, 2589–2607.

Mason, M.G., Mathews, D.E., Argyros, D.A., Maxwell, B.B., Kieber, J.J., Alonso, J.M., Ecker, J.R., and Schaller, G.E. (2005). Multiple type-B response regulators mediate cytokinin signal transduction in Arabidopsis. Plant Cell *17*, 3007–3018.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., et al. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. *34*, D108–D110.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

Megraw, M., and Hatzigeorgiou, A.G. (2010). MicroRNA Promoter Analysis. In Plant MicroRNAs, B.C. Meyers and P.J. Green, eds. (New York: Humana Press), pp. 149–161.

Michael, T.P., Mockler, T.C., Breton, G., McEntee, C., Byer, A., Trout, J.D., Hazen, S.P., Shen, R., Priest, H.D., Sullivan, C.M., et al. (2008). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. PLoS Genet. *4*, e14.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., and Alon, U. (2004). Superfamilies of evolved and designed networks. Science *303*, 1538–1542.

Mittler, R., Kim, Y., Song, L., Coutu, J., Coutu, A., Ciftci-Yilmaz, S., Lee, H., Stevenson, B., and Zhu, J.K. (2006). Gain- and loss-of-function mutations in Zat10 enhance the tolerance of plants to abiotic stress. FEBS Lett. *580*, 6537–6542.

Mok, D.W.S., and Mok, M.C. (1994). CytokininsChemistry, Activity, and Function (Boca Raton: CRC Press).

Neph, S., Stergachis, A.B., Reynolds, A., Sandstrom, R., Borenstein, E., and Stamatoyannopoulos, J.A. (2012a). Circuitry and dynamics of human transcription factor regulatory networks. Cell *150*, 1274–1286.

Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012b). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature *489*, 83–90.

Nover, L., Bharti, K., Döring, P., Mishra, S.K., Ganguli, A., and Scharf, K.D. (2001). Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperones *6*, 177–189.

Pruitt, K.D., Harrow, J., Harte, R.A., Wallin, C., Diekhans, M., Maglott, D.R., Searle, S., Farrell, C.M., Loveland, J.E., Ruef, B.J., et al. (2009). The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. Genome Res. *19*, 1316–1323.

Queitsch, C., Hong, S.W., Vierling, E., and Lindquist, S. (2000). Heat shock protein 101 plays a crucial role in thermotolerance in Arabidopsis. Plant Cell *12*, 479–492.

Rossel, J.B., Wilson, P.B., Hussain, D., Woo, N.S., Gordon, M.J., Mewett, O.P., Howell, K.A., Whelan, J., Kazan, K., and Pogson, B.J. (2007). Systemic and intracellular responses to photooxidative stress in Arabidopsis. Plant Cell *19*, 4091–4110.

Scharf, K.D., Berberich, T., Ebersberger, I., and Nover, L. (2012). The plant heat stress transcription factor (Hsf) family: structure, function and evolution. Biochim. Biophys. Acta *1819*, 104–119.

Schramm, F., Ganguli, A., Kiehlmann, E., Englich, G., Walch, D., and von Koskull-Döring, P. (2006). The heat stress transcription factor HsfA2 serves as a regulatory amplifier of a subset of genes in the heat stress response in Arabidopsis. Plant Mol. Biol. *60*, 759–772.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shin, J., Kim, K., Kang, H., Zulfugarov, I.S., Bae, G., Lee, C.H., Lee, D., and Choi, G. (2009). Phytochromes promote seedling light responses by inhibiting four negatively-acting phytochrome-interacting factors. Proc. Natl. Acad. Sci. USA *106*, 7660–7665.

Sieburth, L.E., and Meyerowitz, E.M. (1997). Molecular dissection of the AGAMOUS control region shows that cis elements for spatial regulation are located intragenically. Plant Cell *9*, 355–365.

Spartz, A.K., Lee, S.H., Wenger, J.P., Gonzalez, N., Itoh, H., Inzé, D., Peer, W.A., Murphy, A.S., Overvoorde, P.J., and Gray, W.M. (2012). The SAUR19 subfamily of SMALL AUXIN UP RNA genes promote cell expansion. Plant J. *70*, 978–990.

Stergachis, A.B., Haugen, E., Shafer, A., Fu, W., Vernot, B., Reynolds, A., Raubitschek, A., Ziegler, S., LeProust, E.M., Akey, J.M., et al. (2013). Exonic transcription factor binding directs codon choice and affects protein evolution. Science *342*, 1367–1372.

Teves, S.S., and Henikoff, S. (2011). Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. Genes Dev. *25*, 2387–2397.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

van Zanten, M., Tessadori, F., Peeters, A.J.M., and Fransz, P. (2012). Shedding light on large-scale chromatin reorganization in Arabidopsis thaliana. Mol. Plant *5*, 583–590.

Vandepoele, K., Vlieghe, K., Florquin, K., Hennig, L., Beemster, G.T., Gruissem, W., Van de Peer, Y., Inzé, D., and De Veylder, L. (2005). Genome-wide identification of potential plant E2F target genes. Plant Physiol. *139*, 316–328.

Weirauch, M., Yang, A., Albu, M., Cote, A., Montenegro-Montero, A., Drewe, P., Najafabadi, H., Lambert, S., Mann, I., Cook, K., Zheng, H., Goity, A., van Bakel, H., Lozano, J., Galli, M., Lewsey, M., Huang, E., Mukherjee, T., Chen, X., Reece-Hoyes, J., Govindarajan, S., Shaulsky, G., Walhout, A.J.M., Bouget, F., Ratsch, G., Larrondo, L., Ecker, J.R., and Hughes, T. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell *158*, Published September 11, 2014. http://dx.doi.org/10.1016/j.cell.2014.08.009.

Wu, C., Wong, Y.-C., and Elgin, S.C.R. (1979). The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. Cell *16*, 807–814.

Yadav, V., Mallappa, C., Gangappa, S.N., Bhatia, S., and Chattopadhyay, S. (2005). A basic helix-loop-helix transcription factor in Arabidopsis, MYC2, acts as a repressor of blue light-mediated photomorphogenic growth. Plant Cell *17*, 1953–1966.

Zhang, Y., Mayba, O., Pfeiffer, A., Shi, H., Tepperman, J.M., Speed, T.P., and Quail, P.H. (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*, e1003244.

Zhong, S., Zhao, M., Shi, T., Shi, H., An, F., Zhao, Q., and Guo, H. (2009). EIN3/EIL1 cooperate with PIF1 to prevent photo-oxidation and to promote greening of Arabidopsis seedlings. Proc. Natl. Acad. Sci. USA *106*, 21431–21436.