Title

A Genome-Scale Resource for the Functional Characterization of Arabidopsis Transcription Factors

Journal

Cell Reports, 8(2)

ISSN

2639-1856

Authors

Pruneda-Paz, Jose L
Breton, Ghislain
Nagel, Dawn H
et al.

Publication Date

2014-07-01

DOI

10.1016/j.celrep.2014.06.033

Peer reviewed

Published in final edited form as:
Cell Rep. 2014 July 24; 8(2): 622–632. doi:10.1016/j.celrep.2014.06.033.

A genome-scale resource for the functional characterization of Arabidopsis transcription factors

Jose L. Pruneda-Paz[1,2,9,*], **Ghislain Breton**[1,4,9], **Dawn H. Nagel**[1,2,7], **S. Earl Kang**[1,2], **Katia Bonaldi**[1,2], **Colleen J. Doherty**[1,2,5], **Stephanie Ravelo**[1,6], **Mary Galli**[3,8], **Joseph R. Ecker**[3], and **Steve A. Kay**[1,2,7,*]

[1]Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093, USA

[2]Center for Chronobiology, University of California San Diego, La Jolla, CA 92093, USA

[3]Genomic Analysis Laboratory, Howard Hughes Medical Institute and The Salk Institute for Biological Studies, La Jolla, CA 92037, USA

SUMMARY

Extensive transcriptional networks play major roles in cellular and organismal functions. Transcript levels are in part determined by the combinatorial and overlapping functions of multiple transcription factors (TFs) bound to gene promoters. TF-promoter interactions thus provide the basic molecular wiring of transcriptional regulatory networks. In plants, discovery of the functional roles of TFs is limited by an increased complexity of network circuitry due to a significant expansion of TF families. Here, we present the construction of a comprehensive clone-collection of Arabidopsis TFs created to provide a versatile resource to uncover TF biological functions. We leveraged this collection by implementing a high-throughput DNA-binding assay and identified direct regulators of a key clock gene (CCA1) that provide molecular links between different signaling modules and the circadian clock. The resources introduced in this work will

[*]Correspondence: jprunedapaz@ucsd.edu (J.L.P-P.); stevekay@usc.edu (S.A.K).
[4]Present address: Department of Integrative Biology and Pharmacology, University of Texas Health Sciences Center-Houston (UTHSC), Houston, TX 77030, USA
[5]Present address: Department of Molecular & Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA
[6]Present address: Division of Rheumatology, Allergy, and Immunology, School of Medicine, University of California San Diego, La Jolla, CA 92093, USA
[7]Present address: Molecular and Computational Biology Section, University of Southern California, Los Angeles, CA 90089, USA
[8]Present address: Waksman Institute of Microbiology, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA
[9]These authors contributed equally to this work.

SUPPLEMENTAL INFORMATION

Supplemental information includes four figures, eight tables, extended experimental procedures and supplemental references, and can be found with this article online at http://

AUTHOR CONTRIBUTIONS

J.L.P-P., G.B., S.E.K., and S.R. generated and validated pENTR-TF and pDEST22-TF clones; J.R.E., and M.G. provided clones initial library construction; J.L.P-P. and G.B. optimized and performed yeast one hybrid screens; K.B. performed protoplast transient assays; C.J.D. performed bioinformatic analyses; J.L.P-P. and DHN performed FBH1 characterization; J.L.P-P. and G.B. compiled and analyzed data for all experiments; J.L.P-P., G.B. and S.A.K. analyzed results and wrote the manuscript; all authors edited the manuscript.

significantly contribute to a better understanding of the transcriptional regulatory landscape of plant genomes.

INTRODUCTION

Transcription factors (TFs) are one of the largest functional class of proteins encoded in eukaryotic genomes, often accounting for almost 8% of the total gene pool (Weirauch and Hughes, 2011). A genome-wide survey of binding sites for 119 human TFs using chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) revealed that even with a partial view of the TF ensemble (119 out of ~2000), the fraction of DNA base-pairs involved in gene regulation is far greater than the protein-coding fraction (ENCODE Project Consortium, 2012). The results from this global survey further confirm that gene expression is regulated by the combinatorial effect of multiple TFs bound to gene regulatory regions. Despite their widespread importance, only a limited number of TFs are characterized at the biochemical and molecular levels.

Pioneering work in sea urchin combining genetics, transcript profiling, and localization of cis-regulatory modules established a gene regulatory network (GRN) model that mapped the regulatory logic of developmental control (Oliveri et al., 2008). Subsequent studies indicate that numerous transcription-based networks are sufficiently hardwired in the genome to be modeled (Davidson, 2006). The construction of circadian clock regulated networks revealed a much more complex circuitry than previously anticipated by forward genetic screens (Koike et al., 2012; Rey et al., 2011; Ueda et al., 2005). Several positive and negative feedback loops appear to have evolved in parallel, which might allow for multiple input signals to provide proper phasing and rhythmic synchrony of biological processes with the oscillating environment (Zhang and Kay, 2010). In addition to this redundant network architecture most plant TF families have significantly expanded during evolution, potentially to diversify the mechanisms necessary to survive the myriad of challenges associated with direct exposure to the environment (Shiu et al., 2005). Thus, network and genetic redundancy likely accounts for a large proportion of circadian and other transcriptional networks' resilience to forward genetic approaches in plants (Pruneda-Paz and Kay, 2010).

To overcome the difficulties associated with network and TF redundancy, novel approaches to physically map cis-regulatory networks have been developed (Bassel et al., 2012). While TF-centered approaches such as ChIP-seq can reveal the extent to which a particular TF is involved in genome-wide regulation, other techniques such as high-throughput yeast one-hybrid (HT-Y1H) are promoter-focused and can provide a survey of potential interactors for a single promoter. Reagents for the latter have been developed and applied successfully to study gene regulation in humans, flies (*Drosophila melanogaster*) and worms (*Caenorhabditis elegans*), and are starting to reveal the complexity of combinatorial gene regulation (Deplancke et al., 2004; Hens et al., 2011; Reece-Hoyes et al., 2011). Recently, we implemented this genomic strategy in plants (*Arabidopsis thaliana*, Arabidopsis) and identified novel transcriptional components of the circadian clock (Pruneda-Paz et al., 2009) as well as other plant physiological processes (Ito et al., 2012; Li et al., 2012; Niwa et al., 2013). In these studies, TF-DNA interactions identified in Y1H screens were confirmed *in*

vivo by ChIP and the biological relevance of identified TFs was demonstrated by reverse genetics approaches. Thus, combined with community-developed reagents such as homozygous insertion line collections (O'Malley and Ecker, 2010) and genome editing strategies (Puchta and Fauser, 2013), HT-Y1H provides a powerful technique to explore the cohort of TFs that bind to any gene promoter. Specifically, large transcriptional regulatory networks, such as the circadian clock, could be comprehensively studied using this approach.

In our previous study, we took advantage of approaches developed for *C. elegans* (Deplancke et al., 2004; Deplancke et al., 2006) and of previously generated Arabidopsis TF collections (Gong et al., 2004; Paz-Ares and Regia, 2002; Underwood et al., 2006; Yamada et al., 2003) to develop HT-Y1H screens for Arabidopsis (Pruneda-Paz et al., 2009). Since our initial arrayed clone collection included the open reading frames (ORFs) for only 186 circadian-regulated TFs (Pruneda-Paz et al., 2009), we decided to generate a genome-wide plasmid library encompassing the ORFs for all known Arabidopsis TFs (TF ORFeome). During the last decade, several efforts were undertaken to annotate Arabidopsis TFs and create digital repositories (Guo et al., 2005; Kummerfeld and Teichmann, 2006; Riano-Pachon et al., 2007; Riechmann et al., 2000), as well as to generate TF clone collections (Gong et al., 2004; Paz-Ares and Regia, 2002; Underwood et al., 2006; Yamada et al., 2003). Building from these resources, we present the construction of the most complete Arabidopsis TF ORFeome available to date. All clones in this collection were individually sequence-validated, have the same reading frame and vector backbone, and are compatible with recombination-based cloning. Thus, quality and versatility are essential features of this ORFeome. This TF collection was used to develop automated HT-Y1H screens that uncovered novel players of the transcriptional networks underlying the Arabidopsis clock function. The resources presented here will empower the plant community with reliable and universal reagents to explore the regulatory landscape of the cis-regulome.

RESULTS

Construction of a TF collection compatible with recombination-based cloning

Structural redundancy inherent to complex positive/negative feedback loops and large expansion of plant TF families greatly hinders our ability to deconvolute the organization of transcription-based regulatory networks such as the Arabidopsis circadian clock. To bypass these limitations, we previously implemented a reverse genetics approach that proved to be successful in other model organisms and identified a TF that regulates the Arabidopsis clock function (Deplancke et al., 2004; Pruneda-Paz et al., 2009). This alternative strategy takes advantage of TF-specific plasmid libraries and the Y1H system (Deplancke et al., 2004). Considering the simplicity and success of the approach and that our first TF collection only included circadian regulated-TFs, we hypothesized that additional clock components could be discovered with a larger gene collection. We therefore aimed at generating a complete sequence-validated Arabidopsis TF library to continue our exploration of the circadian system in a comprehensive and unbiased manner. To achieve this goal, we defined a comprehensive list of Arabidopsis TFs and transcriptional regulators (hereafter globally referred as TFs) by combining all TFs predicted in four independent databases (PlnTFDB,

http://plntfdb.bio.uni-potsdam.de; DATF, http://datf.cbi.pku.edu.cn; DBD, http://www.transcriptionfactor.org and REGIA consortium-personal communication) (Guo et al., 2005; Kummerfeld and Teichmann, 2006; Paz-Ares and Regia, 2002; Riano-Pachon et al., 2007). As each database used alternative TF identification algorithms to survey the Arabidopsis genome, TF gene predictions in each set did not fully overlap. To be as inclusive as possible, all TF encoding genes present in any of these databases (2492 genes) were included in our initial cloning pipeline (Table S1). Grouping TFs by their occurrence in each dataset revealed that approximately an equal proportion of genes are found in one or all databases (Tables S1 and S2). Additionally, based on specific literature searches, a small number of TFs were included (Tables S1 and S2, referred as “other”).

While some Arabidopsis TF ORFeomes were previously generated (Gong et al., 2004; Paz-Ares and Regia, 2002; Underwood et al., 2006; Yamada et al., 2003), our goal was to build a comprehensive, sequence-validated, and homogeneous clone-collection regarding plasmid backbone, resistance marker and coding sequence reading frame. Each TF coding sequence was amplified with its respective STOP codon and cloned in the Gateway compatible vector pENTR/D (Life Technologies). This universal format simplifies bulk downstream usage of the collection to develop TF-centered experimental approaches. Primer pairs for the longest isoform of the 2492 TFs were designed according to TAIR9 gene annotations (http://www.arabidopsis.org) (Lamesch et al., 2012), and used throughout six independent rounds of cloning and sequencing (Fig. 1A). To build upon previous TF collections, we used 1728 clones from the Salk/Stanford/PGEC Consortium, TIGR, REGIA and Yale collections (Gong et al., 2004; Paz-Ares and Regia, 2002; Underwood et al., 2006; Yamada et al., 2003) as PCR templates for the first round. This step yielded 746 pENTR/D-TF clones for the new TF ORFeome (Fig. 1A). The remaining TF coding sequences (1210) were amplified “de novo” using a cDNA pool isolated from one-week old seedlings collected over different times of the day. After the sixth cloning round, we completed a collection of 1956 sequence-confirmed pENTR/D-TF clones representing 78.5 % of all Arabidopsis TFs (Fig. 1A, 1B, 1C, Table S3). Almost all TF genes included in the collection (99 %) encode polypeptides predicted by the current Arabidopsis genomic annotation (TAIR10) (Fig. 1D, Table S4). A fraction of these clones (4.6 %) contain silent point mutations likely generated during gene amplification (Fig. 1D, Tables S3 and S4). The remaining 1 % of the genes included in the collection correspond to coding sequences with conserved mismatches (respect to the current gene annotation) repeatedly observed across most cloning attempts, and therefore likely encode TF isoforms not previously identified (Fig. 1D and S1A, Tables S3 and S4). The collection includes TFs for 93 out of 97 (96 %) predicted Arabidopsis TF families and has a clone-coverage higher than 50 % for 86 (89 %) of these TF families (Fig. 1E and S1B, Table S5). Only 4 TF family singletons (hATP, LFY, NOT, and TBP) are not present in the collection, while other families, such as ARF, PHD and SNF2 are significantly underrepresented (Fig. S1B, Table S5). Cloning these TFs coding sequences seems particularly difficult and may require alternative cloning strategies and/or expression systems.

Throughout the different cloning iterations, each clone status was systematically monitored at different steps allowing us to troubleshoot the cloning procedure for specific groups of genes. For example, cloning coding sequences longer than 2000 bp required an adjustment

of most protocols (details in Extended Experimental Procedures). Out of the 536 missing clones, 135 were not found in any previous collection and never amplified from our cDNA pool, 210 were prone to multiple mutations and never found intact, and the rest could not be cloned for various reasons (Tables S6 and S7). It is worth noting that gene size does not seem to be a limiting factor for our cloning procedure (Table S6). While a small subset of predicted Arabidopsis TFs remain to be cloned, the resource described here represents the largest eukaryotic sequence-validated TF ORF collection available to date. Several genes in our TF ORFeome are either missing (15.1 %) or not specifically detected (3.3 %) by the commonly used ATH1 DNA microarray platform (Affymetrix) (Fig. S1C and S1D). Inclusion of such genes in functional genomic applications will certainly leverage our ability to gain more information and subsequently uncover their functions (individual clones or the entire collection are available through the Arabidopsis Biological Resource Stock Center - www.abrc.osu.edu).

Construction of a TF collection suitable for yeast one-hybrid screens

Tremendous progress has been made in defining the DNA-binding protein landscape using tools such as TF-centered HT-Y1H (Deplancke et al., 2004; Hens et al., 2011; Reece-Hoyes et al., 2011). By taking advantage of the TF collection compatibility with recombination-based cloning (Gateway, Life technologies), we transferred each TF into a Y1H compatible destination plasmid (pDEST22) that carries the GAL4-Activation Domain (GAL4_AD) located 5′ to the TF insertion site. TFs cloned in pDEST22 are thus expressed in yeast as C-terminal fusions to the GAL4_AD. After three rounds of ORF transfer and destination plasmid validation by restriction analysis and gene specific colony PCR, 100% of the 1956 clones were successfully cloned into pDEST22 (Fig. 2A). The pDEST22-TF collection was arrayed in twenty-one 96-well plates. To efficiently manage the larger number of constructs, a newly optimized version of our previous TF library screen protocol was developed (Fig. S2A). Briefly, we generated a modified yeast strain (YU) in order to provide yeast mating capability with the strain YM4271 that carries the *promoter::lacZ* reporter constructs (Fig. S2B). Similar to the procedure that was recently reported (Gaudinier et al., 2011), each pDEST22-TF 96-well plate was transformed into the YU strain using a high-throughput yeast transformation protocol described previously (Walhout and Vidal, 2001). Each plate included one empty well as a negative growth control and one pDEST22 empty plasmid control to set the basal reporter activity in the absence of DNA binding. After transformation, 96-well plates were indexed into six 384-well plates. In this condensed format the YU library strains were mated with promoter-specific YM4271 reporter strains, and the β-galactosidase reporter activity was quantified (Fig. S2A). HT-Y1H screens were initially tested and optimized using a fully automated liquid handling robotic platform (details in Extended Experimental Procedures) and a temporary “bridge” TF library generated while the construction of the final gold standard collection was underway (Li et al., 2012). While automation significantly increases the screening throughput, the overall procedure can be performed manually and thus could be implemented in any laboratory.

Uncovering novel regulators of the *CCA1* promoter activity

In our original TF-centered Y1H screen, the promoter of a key Arabidopsis clock component, *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), was scanned with a library

encompassing 186 circadian-regulated *TFs* (Pruneda-Paz et al., 2009). Since this initial screen uncovered a novel core clock component, called CCA1 HIKING EXPEDITION (CHE), we decided to search for additional *CCA1* promoter regulators with the newly generated pDEST22-TF library. As yeast promoters are rather compact (Dobi and Winston, 2007), instead of screening a long 1–2 kb *CCA1* promoter fragment, we had previously opted to screen five short overlapping promoter segments of ~200 bp in length (Pruneda-Paz et al., 2009). To determine if indeed this strategy increases the sensitivity of HT-Y1H screens, we analyzed the binding of CHE to either the five overlapping *CCA1* promoter fragments or the full-length *CCA1* promoter in a Y1H assay (Fig. S2C). The binding of CHE could only be detected when a short *CCA1* promoter fragment was used indicating that, at least for some TFs such as CHE, the screen sensitivity is significantly improved by reducing the distance between the cis-element and the minimal promoter driving the *lacZ* reporter gene expression (Fig. S2C). Based on this result, the five original *CCA1* promoter fragments were screened with the new TF collection (Fig. 2B). The β-galactosidase reporter activity was determined for each promoter fragment-TF combination and negative controls. Cut-off values were established at four standard deviations (SD) above the average reporter activity obtained for the pDEST22 empty plasmid controls (n=21) (details in Extended Experimental Procedures). Using this criterion, 15 to 160 TF interactions per promoter fragment were identified, for a total of 431 interactions (318 unique TFs) statistically different from the pDEST22 control (Fig. S2D, Table S8). Most of these interactions were specific for a single promoter fragment (73.3 %) but some TFs interacted with two (19.8 %), three (5.3 %), four (1.3 %) or all (0.3 %) promoter baits (Fig. S2E, Table S8). To further reduce the likelihood of false positives, we doubled the cut-off value (8 SD above the average reporter activity obtained for the pDEST22 empty plasmid controls), which resulted in 58 high confidence interactions by 52 unique TFs (Fig. 2C, 2D and 2E). Four of these TFs bind to two or three *CCA1* promoter fragments accounting for 10 high confidence interactions (Fig. 2E). The remaining 48 TFs showed increased β-galactosidase reporter activity with only one *CCA1* promoter bait fragment (Fig. 2E, Table S8). TF family analysis of these interactors revealed a wide array of different families and, in many cases, an enrichment of TFs from the same family on specific fragments (Fig. 2E and 2F). Analysis of the TF family distribution for the 318 TFs found to interact with any of the five *CCA1* promoter fragments indicates a widespread TF family representation (Fig. S2G). In addition, similar TF family binding fragment preferences were found in this larger set of *CCA1* promoter interactors (Fig. S2H, Table S8).

While HT-Y1H screens provide an effective strategy to rapidly identify potential regulators of any given gene promoter, the biological relevance of these regulators must be addressed *in vivo*. CCA1 is a key component of the Arabidopsis clock and its misexpression almost unequivocally results in altered clock function (Nagel and Kay, 2012). Thus, we anticipated that overexpression of novel *CCA1* promoter interactors would perturb the clock function. To perform a rapid validation assessment, Arabidopsis protoplasts were co-transformed with a plasmid carrying a clock reporter construct (*CCA1::LUC+*) and a plasmid carrying an overexpression construct for the potential regulators (*CsVMVS::TF*). After transformation, protoplasts were incubated under constant light conditions and the luciferase activity was monitored every 2.5 h over a period of 5 days. Luciferase traces were used to analyze the

clock function (details in Experimental Procedures) upon overexpression of the top 34 yeast one-hybrid TF candidates (accounting for 38 of the 58 high confidence *CCA1* promoter interactions) (Table S8). Compared to an empty effector plasmid control, the amplitude of recorded luminescence rhythms was significantly reduced when 32 of these TFs were overexpressed, suggesting that all of them could regulate *CCA1* expression (Fig. 3A, Table S8). Rhythmicity profile analysis showed that overexpression of 21 out of the 32 TFs exhibiting amplitude effects also resulted in arrhythmic *CCA1* promoter activity (Fig. 3B, 3C and S3A; Table S8). Our results indicate that top TF candidates tend to result in stronger (i.e. arrhythmic) clock phenotypes, which suggests that HT-Y1H candidate prioritization based on the strength of the β-galactosidase reporter activity provides an effective mean to predict biologically relevant interactions (Fig. 3D). Analysis of microarray datasets did not reveal strong expression correlation or anti-correlation between *CCA1* and most protoplast-validated *TF* (Fig. 3E and S3B). This result indicates that *CCA1* regulators are not necessarily co-expressed with *CCA1* thus emphasizing the advantage of using HT-Y1H for TF discovery. The circadian clock is postulated to be a highly interconnected regulatory hub regulated by multiple environmental and endogenous signals (Pruneda-Paz and Kay, 2010). To gain insights on the possible biological role of these TFs we wanted to examine their potential functional associations by performing Gene Ontology (GO) analysis. However, since little is known about the function of these TFs, no GO annotation other than their classification as transcriptional regulators existed for the majority of them. Therefore, for the 28 protoplast-validated TFs that are specifically detected on the ATH1 arrays (Affymetrix), we searched for first order interaction partners in a publicly available Arabidopsis protein interaction dataset (http://interactome.dfci.harvard.edu/A_thaliana/) (Arabidopsis Interactome Mapping, 2011) (Fig. S3C). This analysis revealed interaction networks, for 10 out of the 28 protoplast-validated TFs, significantly enriched for a wide-range of GO biological process terms (Fig. 3F), consistent with the expansive role of the circadian clock. Altogether, these results suggest that a multitude of TFs likely mediate the regulation of *CCA1* by different endogenous and environmental signals.

Characterization of the regulation of Arabidopsis clock function by FBH1

A recently characterized TF named FLOWERING BHLH 1 (FBH1) (AT1G35460) (Ito et al., 2012) was among the TFs that display clock phenotypes in transient Arabidopsis protoplast assays (Fig. 3A, 3B, 3C and S3A). Characterization of this TF has been likely limited by the fact that *FBH1*-associated probe sets on the ATH1 array exhibit non-specific cross hybridization (Table S1). For this reason, we decided to further explore the function of FBH1 on the regulation of *CCA1*. Y1H assays performed in a 96-well format confirmed that FBH1 binds to the -213/-42 region of the *CCA1* promoter (Fig. 4A). Two potential FBH1 binding sites, a canonical E-box motif (CANNTG) and a non-canonical one (CANNNG), map to this region of the *CCA1* promoter (Fig. 4B). To determine the functionality of these motifs, mutations were introduced and Y1H reporter strains carrying these mutations in the context of the -213/-42 region of the *CCA1* promoter were generated (Fig. 4B). Reporter activation indicates that FBH1 binds preferentially to the non-canonical E-box like motif CACTAG (Fig. 4B and 4C). To analyze the binding of FBH1 to the *CCA1* promoter *in planta*, chromatin immunoprecipitation (ChIP) experiments were performed using green fluorescent protein-tagged (GFP) *FBH1* overexpression lines. The same clock phenotypes

were observed when tagged or untagged versions of FBH1 were overexpressed, suggesting that GFP-FBH1 retains its function (S4D, S4E and S4F). ChIP results confirmed that FBH1 specifically binds to the *CCA1* promoter *in vivo* (Fig. 4D).To determine FBH1-associated clock phenotypes *in planta*, homozygous *FBH1*-overexpression/*CCA1::LUC+* Arabidopsis lines were obtained (Fig. S4A). Bioluminescence time-course expression analysis using these Arabidopsis lines indicate that the overall *CCA1* promoter activity is significantly reduced suggesting that FBH1 is a repressor of *CCA1* (Fig. 4E). In addition, while the period of clock controlled oscillations remains unaltered in these lines, the phase of *CCA1* peak exhibits ~1h advance (Fig. 4E, 4F, 4G, S4B and S4C). Altogether these results indicate that FBH1 is a novel clock component that directly and negatively regulates *CCA1*. Similar experiments will be required to characterize the function of the remaining *CCA1* regulators identified in this work.

DISCUSSION

By combining community-based gene annotation, previous ORFeome efforts and our optimized pipeline for gene cloning and transfer, we created the most comprehensive eukaryotic TF ORF clone collection available to date (1956 TFs). This collection is compatible with recombination-based cloning, which allows bulk transfer of TF coding sequences into any destination plasmid such as the ones already created by the Arabidopsis community (Karimi et al., 2002; Rhee et al., 2003). The collection was conceived as a high-throughput resource to generate large genomic tools for the study of Arabidopsis TF function (Castrillo et al., 2011; Chang et al., 2013; Ou et al., 2011; Siggers et al., 2012; Wehner et al., 2011). One of the main goals of our project was to provide the highest clone quality. Indeed, the nucleotide sequence for 1847 of the 1956 clones included in the collection is identical to the current gene annotation (TAIR10). Despite our effort, the collection is still missing 536 clones. A survey of our cloning pipeline history indicates that 25% of these genes were never detected by PCR (Table S6). These might correspond to pseudo-genes or genes not present in the cDNA pool used as PCR template. They could also be mis-annotated genes not amplifiable with the PCR primer set used. In addition, most of the remaining missing genes were successfully PCR amplified but never found as an intact or correct clone. It is possible that toxicity in *E. coli*, assuming that the cloned gene is translated, could account for this outcome. This seems to be the case, as almost 65% of these clones had deleterious point mutations, were consistently truncated at different locations or could never be found at all (Table S6). Future attempts to complete the collection will need to consider alternative bacterial hosts or cloning and maintaining these constructs directly in yeast cells.

By following an iterative procedure like the one used to obtain the pENTR/D-TF clones, and applying a stringent quality control after recombination based cloning, all TF coding sequences were successfully transferred into a Y1H compatible plasmid (pDEST22). This achievement allowed us to generate the most comprehensive TF collection for promoter DNA-binding protein profiling. While most of previous Y1H screens in eukaryotes were performed using large promoter fragments, we found that the use of shorter overlapping fragments provides increased sensitivity to the assay (Fig. S2C). Such experimental design revealed a total of 318 TFs that account for 431 interactions to the *CCA1* promoter region

used in the screen. While a similar proportion of these interactions were found to bind to most promoter baits (68 to 160), a significantly lower number (15 TFs) was found for one particular promoter fragment (-213/-42). This is likely due to the elevated basal β-galactosidase activity detected with this reporter strain, which only allows the detection of TFs that strongly induce the expression of the *lacZ* reporter. How many TFs can interact *in vivo* with a particular promoter? Additional large-scale deep screening of specific promoter fragments using HT-Y1H are likely to provide this answer in the near future. Recent results from comprehensive monitoring of DNA binding for more than one hundred TFs by ChIP-seq, indicate that promoters are bound by several different TFs, each of which binds to thousands of sites in the genome. This survey suggests that complex combinatorial gene regulation may be the norm in eukaryotes (Consortium et al., 2012). This is in-line to the concept of “billboard” enhancers developed from the study of fly enhancers, in which a flexible group of diverse DNA binding proteins that aggregate depending on specific conditions suffices to provide time and condition dependent gene regulation (Arnosti and Kulkarni, 2005). In such a scenario, we could imagine that promoters containing hundreds of different TF binding sites might, depending on the tissue, developmental stage or time of day, help to coalesce a specific combination of TFs to support a specific biological function. It is certainly premature to provide definitive answers but tools such as the TF ORFeome presented here will certainly help in advancing towards this goal.

By clustering the 58 high confidence *CCA1* promoter interactors by TF family, we observe that some of them display a significant binding preference for a specific *CCA1* promoter fragment (Fig 2E). This trend is maintained across all 318 interactors in our study (Table S8). For example, most interactions by G2-like (63%) TFs are detected with the +88/+259 *CCA1* promoter bait strain (Fig. S2H). Similarly, 53 % of all TCP interactions are detected with the -363/-192 CCA1 promoter bait strain (Fig. S2H). Most of these TCPs (8 out of 10) belong to the same clade, class I TCP TFs, suggesting that binding specificity is driven by the conserved DNA-binding domain among these TFs, and the presence of a specific cis-element in the -363/-192 *CCA1* promoter region. In fact, we have recently characterized the binding of one of these TFs, named CHE, to a canonical TCP class I-binding site (GGNCCCAC) centered at nucleotides -227/-228 of the *CCA1* promoter (Fig. S3B) (Pruneda-Paz et al., 2009). Our results also indicate that some TF families display low promoter fragment specificity (Fig 2F). For example, 6 GeBP TFs account for 20 interactions (3.3 interactions per TF) that include all the *CCA1* promoter fragments tested (Fig. S2H). This interaction promiscuity is likely due to low DNA binding affinity, broad binding preference and/or lack of binding partners as we did not identify common 5–10 nucleotide motifs between two or more *CCA1* promoter fragments (data not shown). On the other hand, 20 AP2 TFs were also bound to all but one of the *CCA1* promoter fragments tested (Fig. S2H). However, in this case these TFs account for only 28 interactions (1.4 interactions per TF), suggesting that these AP2 TFs likely share a high DNA-binding specificity, but across the family, different clades may recognize alternative cis-elements. Some TF families are not present in our interactor dataset suggesting that the *CCA1* promoter regions tested in this work do not encode the corresponding TF binding site. This seems to be the case for ARF TFs, as the canonical AUX-RE (TGTCTC) is not present in the analyzed *CCA1* promoter fragments. However, it is also possible that the binding of

some TFs cannot be detected either because they are not properly expressed in yeast, or require plant-specific interaction partners or post-translational modifications. Altogether, these results indicate that in addition to identifying TFs, HT-Y1H provides novel insights to understanding TF's DNA binding modes. Exploration of genome-wide TF binding preferences as well as TF structural and phylogenetic analyses will be required in order to further assess these observations (Badis et al., 2009; Weirauch and Hughes, 2011).

Secondary testing of top priority HT-Y1H hits in a transient protoplast system revealed that most interactors could perturb *CCA1* expression and the overall clock function *in vivo* (Fig 3A and 3B). Interestingly, most of these TFs do not share a high expression correlation with *CCA1* and are associated to a wide range of biological functions (Fig. 3E and 3F). These results suggest that a multitude of input signals are necessary for proper clock regulation and converge on the regulation of *CCA1*. This is in line with the critical role of CCA1 in resetting the clock and a number of recent reports indicating that the circadian clock function is modulated by many signals in addition to light and temperature (Haydon et al., 2013; Hong et al., 2013; Pruneda-Paz and Kay, 2010). The increased regulatory complexity suggested by our results would improve network robustness while providing fine-tuning capability and ultimately allow optimal synchronization of endogenous rhythms with the periodic changes in external environmental conditions. Further characterization in Arabidopsis seedlings of one of the TFs identified in our HT-Y1H screen, FBH1, confirmed that this is a novel regulator of *CCA1*. FBH1 binds preferentially to an E-box like motif (CACTAG) centered 405/406 nucleotides upstream of the coding sequence start site. This motif contains only one mismatch with the recently published E-box motif (CACTTG) bound by FBH1 on the promoter of the flowering regulator CONSTANS (Ito et al., 2012). Interestingly, our screen did not indicate any binding activity for FBH2 (AT4G09180), a homolog of FBH1, suggesting that the E-box like motif found in the *CCA1* promoter may be preferentially targeted by FBH1. Alternatively, the binding of FBH1/FBH2 may be enhanced by the presence of a cluster containing three tandem E-box motifs within the *CO* promoter. While FBH1 was described as a *CO* activator, our experiments suggest that it functions as a repressor of *CCA1*. It is possible that the different promoter contexts at the *CCA1* and *CO loci* are responsible for this opposing response (Gordan et al., 2013). Screening the *CO* promoter with the expanded TF ORFeome presented here could reveal insightful differences between the potential regulators for both genes. Interestingly, the clock phase advance observed in *FBH1* overexpression lines is consistent with a similar phenotype observed in overexpression lines for a recently described *CCA1* repressor (i.e. CHE) (Pruneda-Paz et al., 2009). While more work is needed to further confirm the potential *CCA1* promoter interactome suggested by the experiments in Arabidopsis protoplasts, the overall results suggest that multiple plant-signaling modules fine-tune the clock function through some of the TFs uncovered in this work.

Regulation of gene expression relies on the coordinated interaction of specific combinations of TFs bound to gene promoter regions. Globally, all TF-DNA interactions determine the basic architecture of gene regulatory networks controlling fundamental aspects of cellular and organismal functions. This work introduces a valuable community resource to investigate the regulatory landscape of any Arabidopsis gene. Moreover, it provides the

foundation to develop novel approaches that will ultimately reveal the molecular circuitry regulating plant transcriptomes.

EXPERIMENTAL PROCEDURES

Details are available in Extended Experimental Procedures

Construction of a gateway compatible Arabidopsis TF collection

Arabidopsis TF coding sequences were PCR amplified and cloned in the pENTR/D Gateway donor vector (Life Technologies). All clones included in the collection (1956) were sequence-validated (supplemental table 2).

Yeast one-hybrid screens using a genome-wide TF collection

To generate a TF collection suitable for yeast one-hybrid screens, all TF coding sequences were transferred from pENTR/D to pDEST22 (Life Technologies) via Gateway recombination-based cloning (Life Technologies). In pDEST22-TF clones, the GAL4-activation domain is fused to the N-terminus of the TF. The pDEST22-TF library was arrayed in a 96-well format and transformed into the YU yeast strain (generated in this work) as previously described (Walhout and Vidal, 2001). For yeast one-hybrid screens, the resulting yeast collection was indexed into 384-well plates. In this format, YU yeast TF strains (MATα) were mated with YM4271 yeast strains (MATa) carrying chromosomally integrated *CCA1::lacZ* reporter constructs. After mating, diploid cells were enriched and the β-galactosidase activity for each well was determined as described by the manufacturer (Clontech) but with modifications that allowed the assay to be performed in 384-well plates using a robotic platform (Biocel1200, Agilent Technologies). β-galactosidase activities were normalized to the average value obtained for control wells where the pDEST22-TF plasmid was replaced by the pEXP-AD (empty pDEST22) control plasmid (control-normalized). Binding cut-off values for each *CCA1* promoter fragment were set at 4 and 8 standard deviations above the mean control-normalized value obtained for all control wells. To compare results across all *CCA1* promoter fragment screens, control-normalized values were normalized to the cut-off value for each *CCA1* promoter fragment dataset (cut off-normalized). A compiled list of all *CCA1* promoter-interacting candidates was generated using cut-off-normalized values (Table S8).

Arabidopsis protoplasts isolation and transformation

Candidate TF coding sequences were transferred from pENTR/D to the TF-overexpression vector pCsVMV-GW generated in this work using the pCsVMV-PP2C-AmiR vector backbone (Kim and Somers, 2010) via Gateway recombination-based cloning (Life Technologies). Arabidopsis protoplasts were isolated from 4–5 weeks old Columbia ecotype (Col-0) seedlings using a procedure adapted from previously published protocols (Kim and Somers, 2010; Wu et al., 2009), and aliquoted in white 96-well plates (Evergreen Scientific). Protoplasts were co-transformed following a previously published procedure (Wehner et al., 2011), with a TF-overexpression (or control) vector (5 μg per 4 kb) and the reporter vector pOmegaCCA1-LUC_SK+ (Kim and Somers, 2010) (1 μg per 4 kb) that contains a *CCA1::LUC+* reporter construct. After plasmid co-transformation, cells were resuspended

in a solution containing 5% fetal bovine serum (Sigma) and 50 μM luciferin (Biosynth) as described before (Kim and Somers, 2010), and 96-well plates were covered with a transparent plastic lid. Bioluminescence imaging was performed as described below.

Plant materials and growth conditions

Arabidopsis thaliana (Arabidopsis) seedlings used in this work were from the Columbia ecotype (Col-0). The CCA1::LUC+ reporter line was previously described (Pruneda-Paz et al., 2009). To generate *FBH1* overexpression lines, the pENTR/D-FBH1 construct (plate U21 well D01, supplemental table 2) was used to transfer *FBH1* coding sequence into the pB7WG2 binary vector (Karimi et al., 2002) via Gateway recombination-based cloning (Life Technologies). Similarly, the FBH1-YFP tag overexpression vector pE104-FBH1, was created by transferring FBH1 coding sequence from pENTR/D-FBH1 into the pEarley104 binary vector (Earley et al., 2006). These binary vectors were transferred into the Arabidopsis CCA1::LUC+ line by *Agrobacterium*-mediated transformation, as described before (Zhang et al., 2006). Unless otherwise stated, plants were grown in Murashige-Skoog medium (1.5% agar) supplemented with 3% sucrose under 12 hour light (70 $\mu mol.m^{-2}.s^{-1}$)/12 hour dark cycles (LD) at 22°C. Chromatin immunoprecipitation assays were performed as previously described (Pruneda-Paz et al., 2009).

Bioluminescence detection and data analysis

Seeds were stratified for 2–3 days at 4°C and grown in Murashige-Skoog medium (1.5% agar) supplemented with 3% sucrose under 12 hour light (70 $\mu mol.m^{-2}.s^{-1}$)/12 hour dark cycles (LD) at 22°C. After 7 days, plates were transferred to constant light (70 $\mu mol.m^{-2}.s^{-1}$) (LL), sprayed with 1 mM luciferin (Biosynth), and the emitted luminescence was analyzed every 2.5 hours during 5 days using a digital CCD camera (Hamamatsu). Similarly, transformed Arabidopsis protoplasts were incubated in LD for 36 hours (two dark and one light periods), transferred to LL and imaged as described above. Images were processed using Metamorph imaging software (Molecular Devices), and data analyzed by Fast Fourier Transform-Non Linear Least Squares (FFT-NLLS) (Plautz et al., 1997) using the interface provided by the Biological Rhythms Analysis Software System (BRASS) (available at http://www.amillar.org) (Southern et al., 2006).

Resource distribution

Both clone collections, pENTR/D-TF and pDEST22-TF, are available at the Arabidopsis Biological Resource Center (ABRC) (http://abrc.osu.edu).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Elsebeth Kolmos, Marian Nohales and Sabrina Sanchez for critical reading of the manuscript; Robert Schmitz for helpful discussions and technical tips; Dr. David Somers for providing the pOmegaCCA1-LUC_SK+, pCsVMV-PP2C-AmiR and pCsVMV-AmiR plasmids; Kim Jeongsik for advice on the protoplast isolation protocol; Leann Ortmann and Jasmine King for technical assistance. J.R.E is a Howard Hughes Medical Institute and Gordon and Betty Moore Investigator (GMBF3034).

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award numbers R01GM056006, R01GM067837, and RC2GM092412 to S.A.K; and R01GM056006 to J.L.P-P. as co-investigator.

References

Arabidopsis Interactome Mapping C. Evidence for network evolution in an Arabidopsis interactome map. Science. 2011; 333:601–607. [PubMed: 21798944]

Arnosti DN, Kulkarni MM. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? Journal of cellular biochemistry. 2005; 94:890–898. [PubMed: 15696541]

Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, et al. Diversity and complexity in DNA recognition by transcription factors. Science. 2009; 324:1720–1723. [PubMed: 19443739]

Bassel GW, Gaudinier A, Brady SM, Hennig L, Rhee SY, De Smet I. Systems analysis of plant functional, transcriptional, physical interaction, and metabolic networks. The Plant cell. 2012; 24:3859–3875. [PubMed: 23110892]

Castrillo G, Turck F, Leveugle M, Lecharny A, Carbonero P, Coupland G, Paz-Ares J, Onate-Sanchez L. Speeding cis-trans regulation discovery by phylogenomic analyses coupled with screenings of an arrayed library of Arabidopsis transcription factors. PloS one. 2011; 6:e21524. [PubMed: 21738689]

Chang KN, Zhong S, Weirauch MT, Hon G, Pelizzola M, Li H, Huang SS, Schmitz RJ, Urich MA, Kuo D, et al. Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. eLife. 2013; 2:e00675. [PubMed: 23795294]

Davidson, EH. The regulatory genome : gene regulatory networks in development and evolution. Oxford: Academic Press; 2006.

Deplancke B, Dupuy D, Vidal M, Walhout AJ. A gateway-compatible yeast one-hybrid system. Genome research. 2004; 14:2093–2101. [PubMed: 15489331]

Deplancke B, Mukhopadhyay A, Ao W, Elewa AM, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Reece-Hoyes JS, Hope IA, et al. A gene-centered C. elegans protein-DNA interaction network. Cell. 2006; 125:1193–1205. [PubMed: 16777607]

Dobi KC, Winston F. Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. Molecular and cellular biology. 2007; 27:5575–5586. [PubMed: 17526727]

Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, Pikaard CS. Gateway-compatible vectors for plant functional genomics and proteomics. The Plant journal : for cell and molecular biology. 2006; 45:616–629. [PubMed: 16441352]

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]

Gaudinier A, Zhang L, Reece-Hoyes JS, Taylor-Teeples M, Pu L, Liu Z, Breton G, Pruneda-Paz JL, Kim D, Kay SA, et al. Enhanced Y1H assays for Arabidopsis. Nature methods. 2011; 8:1053–1055. [PubMed: 22037706]

Gong W, Shen YP, Ma LG, Pan Y, Du YL, Wang DH, Yang JY, Hu LD, Liu XF, Dong CX, et al. Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. Plant physiology. 2004; 135:773–782. [PubMed: 15208423]

Gordan R, Shen N, Dror I, Zhou T, Horton J, Rohs R, Bulyk ML. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. Cell reports. 2013; 3:1093–1104. [PubMed: 23562153]

Guo A, He K, Liu D, Bai S, Gu X, Wei L, Luo J. DATF: a database of Arabidopsis transcription factors. Bioinformatics. 2005; 21:2568–2569. [PubMed: 15731212]

Haydon MJ, Mielczarek O, Robertson FC, Hubbard KE, Webb AA. Photosynthetic entrainment of the Arabidopsis thaliana circadian clock. Nature. 2013; 502:689–692. [PubMed: 24153186]

Hens K, Feuz JD, Isakova A, Iagovitina A, Massouras A, Bryois J, Callaerts P, Celniker SE, Deplancke B. Automated protein-DNA interaction screening of Drosophila regulatory elements. Nature methods. 2011; 8:1065–1070. [PubMed: 22037703]

Hong S, Kim SA, Guerinot ML, McClung CR. Reciprocal interaction of the circadian clock with the iron homeostasis network in Arabidopsis. Plant physiology. 2013; 161:893–903. [PubMed: 23250624]

Ito S, Song YH, Josephson-Day AR, Miller RJ, Breton G, Olmstead RG, Imaizumi T. FLOWERING BHLH transcriptional activators control expression of the photoperiodic flowering regulator CONSTANS in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:3582–3587. [PubMed: 22334645]

Karimi M, Inze D, Depicker A. GATEWAY vectors for Agrobacterium-mediated plant transformation. Trends in plant science. 2002; 7:193–195. [PubMed: 11992820]

Kim J, Somers DE. Rapid assessment of gene function in the circadian clock using artificial microRNA in Arabidopsis mesophyll protoplasts. Plant physiology. 2010; 154:611–621. [PubMed: 20709829]

Koike N, Yoo SH, Huang HC, Kumar V, Lee C, Kim TK, Takahashi JS. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. Science. 2012; 338:349–354. [PubMed: 22936566]

Kummerfeld SK, Teichmann SA. DBD: a transcription factor prediction database. Nucleic acids research. 2006; 34:D74–81. [PubMed: 16381970]

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic acids research. 2012; 40:D1202–1210. [PubMed: 22140109]

Li L, Ljung K, Breton G, Schmitz RJ, Pruneda-Paz J, Cowing-Zitron C, Cole BJ, Ivans LJ, Pedmale UV, Jung HS, et al. Linking photoreceptor excitation to changes in plant architecture. Genes & development. 2012; 26:785–790. [PubMed: 22508725]

Nagel DH, Kay SA. Complexity in the wiring and regulation of plant circadian networks. Current biology : CB. 2012; 22:R648–657. [PubMed: 22917516]

Niwa M, Daimon Y, Kurotani K, Higo A, Pruneda-Paz JL, Breton G, Mitsuda N, Kay SA, Ohme-Takagi M, Endo M, et al. BRANCHED1 interacts with FLOWERING LOCUS T to repress the floral transition of the axillary meristems in Arabidopsis. The Plant cell. 2013; 25:1228–1242. [PubMed: 23613197]

O'Malley RC, Ecker JR. Linking genotype to phenotype using the Arabidopsis unimutant collection. The Plant journal : for cell and molecular biology. 2010; 61:928–940. [PubMed: 20409268]

Oliveri P, Tu Q, Davidson EH. Global regulatory logic for specification of an embryonic cell lineage. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:5955–5962. [PubMed: 18413610]

Ou B, Yin KQ, Liu SN, Yang Y, Gu T, Wing Hui JM, Zhang L, Miao J, Kondou Y, Matsui M, et al. A high-throughput screening system for Arabidopsis transcription factors and its application to Med25-dependent transcriptional regulation. Molecular plant. 2011; 4:546–555. [PubMed: 21343311]

Paz-Ares J, Regia C. REGIA, an EU project on functional genomics of transcription factors from Arabidopsis Thaliana. Comparative and functional genomics. 2002; 3:102–108. [PubMed: 18628849]

Plautz JD, Straume M, Stanewsky R, Jamison CF, Brandes C, Dowse HB, Hall JC, Kay SA. Quantitative analysis of Drosophila period gene transcription in living animals. Journal of biological rhythms. 1997; 12:204–217. [PubMed: 9181432]

Pruneda-Paz JL, Breton G, Para A, Kay SA. A functional genomics approach reveals CHE as a component of the Arabidopsis circadian clock. Science. 2009; 323:1481–1485. [PubMed: 19286557]

Pruneda-Paz JL, Kay SA. An expanding universe of circadian networks in higher plants. Trends in plant science. 2010; 15:259–265. [PubMed: 20382065]

Puchta H, Fauser F. Synthetic nucleases for genome engineering in plants: prospects for a bright future. The Plant journal : for cell and molecular biology. 2013

Reece-Hoyes JS, Barutcu AR, McCord RP, Jeong JS, Jiang L, MacWilliams A, Yang X, Salehi-Ashtiani K, Hill DE, Blackshaw S, et al. Yeast one-hybrid assays for gene-centered human gene regulatory network mapping. Nature methods. 2011; 8:1050–1052. [PubMed: 22037702]

Rey G, Cesbron F, Rougemont J, Reinke H, Brunner M, Naef F. Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. PLoS biology. 2011; 9:e1000595. [PubMed: 21364973]

Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, et al. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic acids research. 2003; 31:224–228. [PubMed: 12519987]

Riano-Pachon DM, Ruzicic S, Dreyer I, Mueller-Roeber B. PlnTFDB: an integrative plant transcription factor database. BMC bioinformatics. 2007; 8:42. [PubMed: 17286856]

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science. 2000; 290:2105–2110. [PubMed: 11118137]

Shiu SH, Shih MC, Li WH. Transcription factor families have much higher expansion rates in plants than in animals. Plant physiology. 2005; 139:18–26. [PubMed: 16166257]

Siggers T, Chang AB, Teixeira A, Wong D, Williams KJ, Ahmed B, Ragoussis J, Udalova IA, Smale ST, Bulyk ML. Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-kappaB family DNA binding. Nature immunology. 2012; 13:95–102. [PubMed: 22101729]

Southern MM, Brown PE, Hall A. Luciferases as reporter genes. Methods in molecular biology. 2006; 323:293–305. [PubMed: 16739586]

Ueda HR, Hayashi S, Chen W, Sano M, Machida M, Shigeyoshi Y, Iino M, Hashimoto S. System-level identification of transcriptional circuits underlying mammalian circadian clocks. Nature genetics. 2005; 37:187–192. [PubMed: 15665827]

Underwood BA, Vanderhaeghen R, Whitford R, Town CD, Hilson P. Simultaneous high-throughput recombinational cloning of open reading frames in closed and open configurations. Plant biotechnology journal. 2006; 4:317–324. [PubMed: 17147637]

Walhout AJ, Vidal M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. Methods. 2001; 24:297–306. [PubMed: 11403578]

Wehner N, Weiste C, Droge-Laser W. Molecular screening tools to study Arabidopsis transcription factors. Frontiers in plant science. 2011; 2:68. [PubMed: 22645547]

Weirauch MT, Hughes TR. A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. Sub-cellular biochemistry. 2011; 52:25–73. [PubMed: 21557078]

Wu FH, Shen SC, Lee LY, Lee SH, Chan MT, Lin CS. Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. Plant methods. 2009; 5:16. [PubMed: 19930690]

Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, et al. Empirical analysis of transcriptional activity in the Arabidopsis genome. Science. 2003; 302:842–846. [PubMed: 14593172]

Zhang EE, Kay SA. Clocks not winding down: unravelling circadian networks. Nature reviews Molecular cell biology. 2010; 11:764–776.

Zhang X, Henriques R, Lin SS, Niu QW, Chua NH. Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nature protocols. 2006; 1:641–646.

HIGHLIGHTS

- Versatile Arabidopsis TF clone-collection compatible with recombination-based cloning.
- High-throughput TF-centered yeast one-hybrid screens.
- Discovery of novel transcriptional components of the Arabidopsis circadian clock.

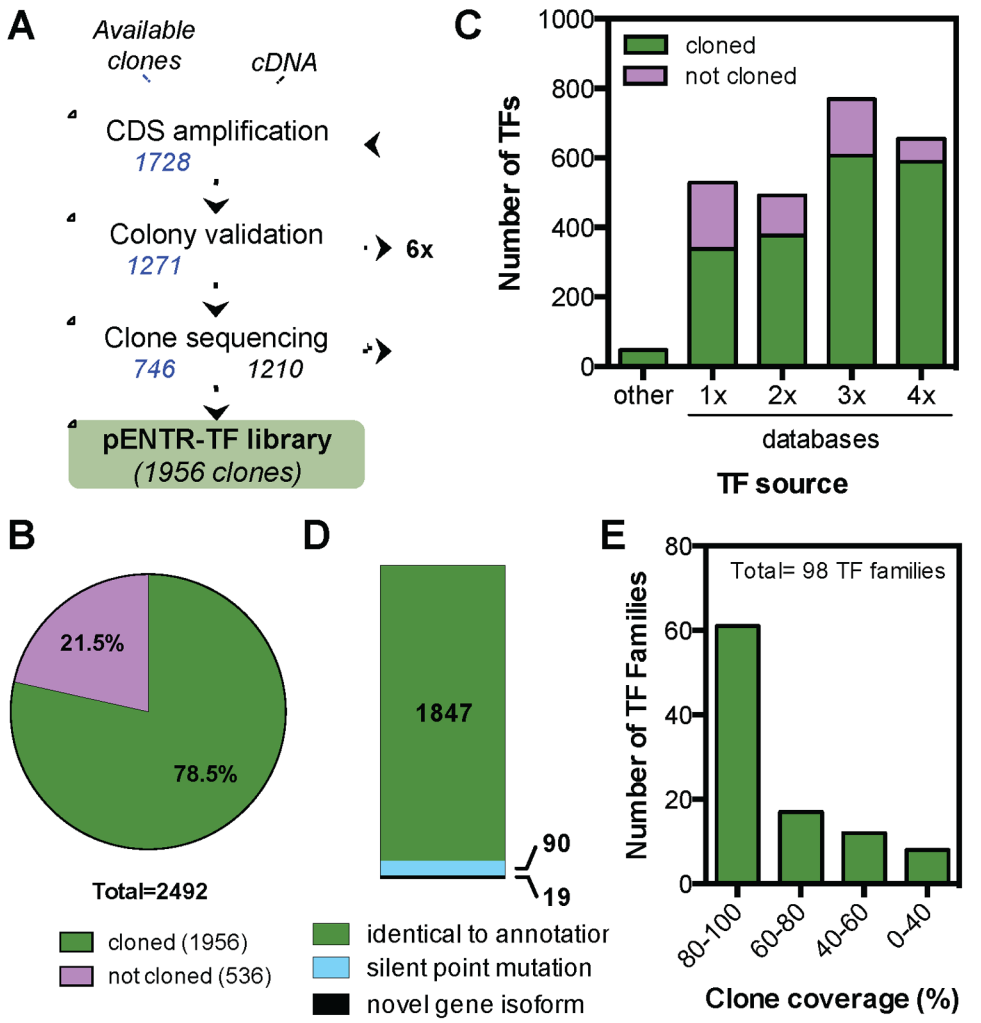

Figure 1. Construction of a genome-wide Arabidopsis TF collection

A) Cloning workflow. Each TF coding sequence was amplified from available clones or "de novo" from cDNA (the number of coding sequences obtained at each step is indicated in blue or red respectively). **B)** Coverage of sequence-validated clones included in the collection. **C)** Frequency distribution of Arabidopsis TFs in four independent databases (cloned genes in each category is indicated in green) (Table S1). **D and E)** Sequence quality (D) and TF family coverage (E) for the clones included in the collection.

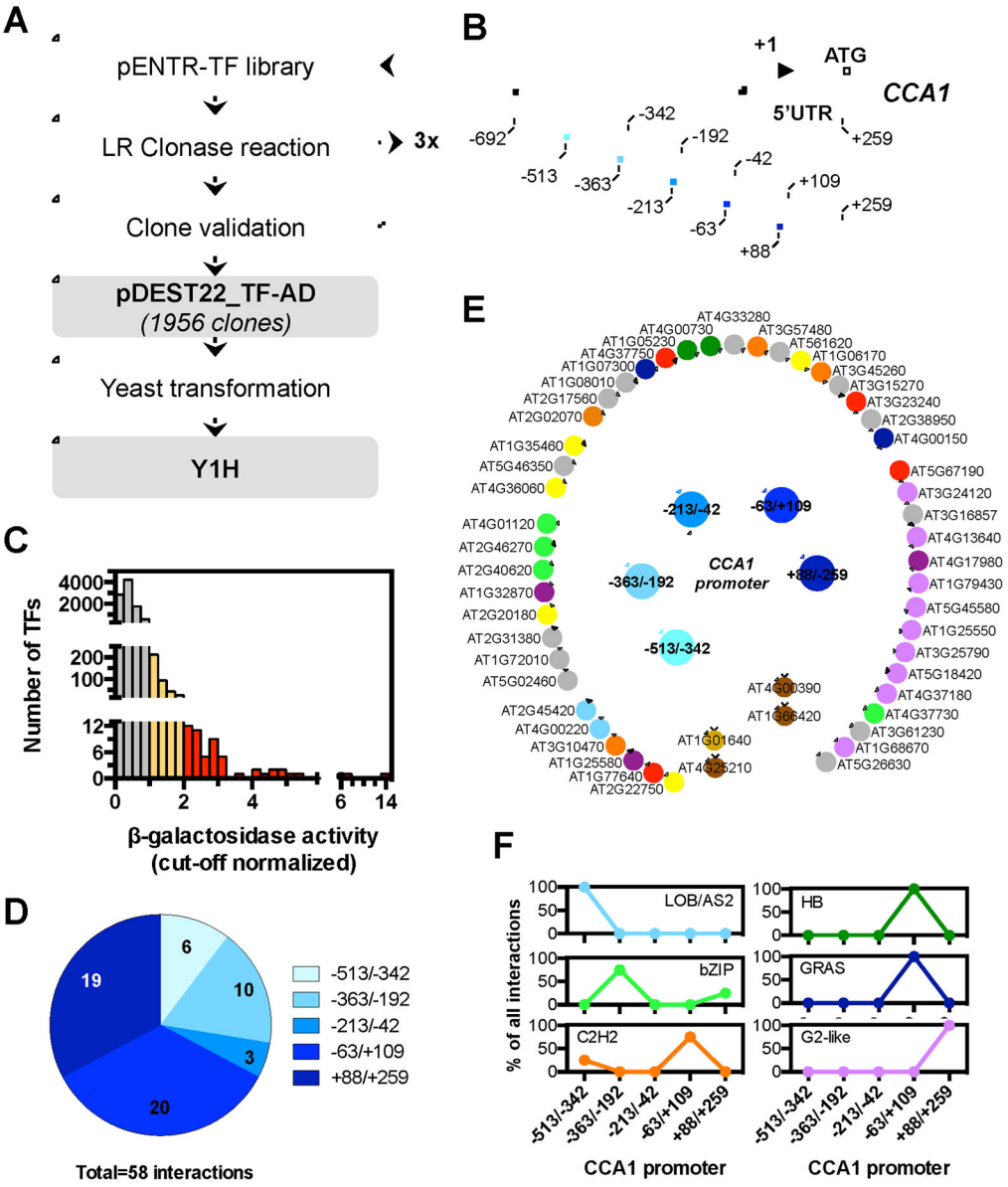

Figure 2. *CCA1* promoter yeast one-hybrid screens using a genome-wide Arabidopsis TF collection

A) Workflow for the construction and validation of a "daughter" pDEST22-TF collection suitable for HT-Y1H screens. **B)** Schematic of the screened *CCA1* promoter fragment baits. **C)** Histogram of the HT-Y1H results obtained for all *CCA1* promoter fragments (Table S8). β-galactosidase activities obtained for each TF-promoter combination were normalized to the cut-off value established for each promoter fragment. Bars represent the number of TF-promoter combinations that resulted in cut-off normalized values below (grey) or above (orange and red) the cut-off value (binwidth=0.25). Red bars indicate high confidence interactions. **D)** Distribution of high confidence TF interaction hits for each *CCA1* promoter fragment (Table S8). **E)** *CCA1* promoter TF interaction network. Interacting TFs and the corresponding Arabidopsis gene identification (AGI) numbers are organized clockwise

based on the β-galactosidase reporter activity obtained in the yeast one-hybrid screen. Interactions by a single member of a TF family (grey) or multiple members of the bHLH (yellow), AP2-ERBP (red), NAC (purple), C2H2 (orange), LOB/AS2 (light blue), bZIP (light green), GRAS (blue), HB (green), G2-like (light purple), TRAF (light brown), or GeBP (brown) TF families are indicated (Table S8). **F**) Frequency distribution for the binding of multiple TF family members to each of the *CCA1* promoter fragments used in the HT-Y1H screen. TF families with high promoter fragment specificity are shown.

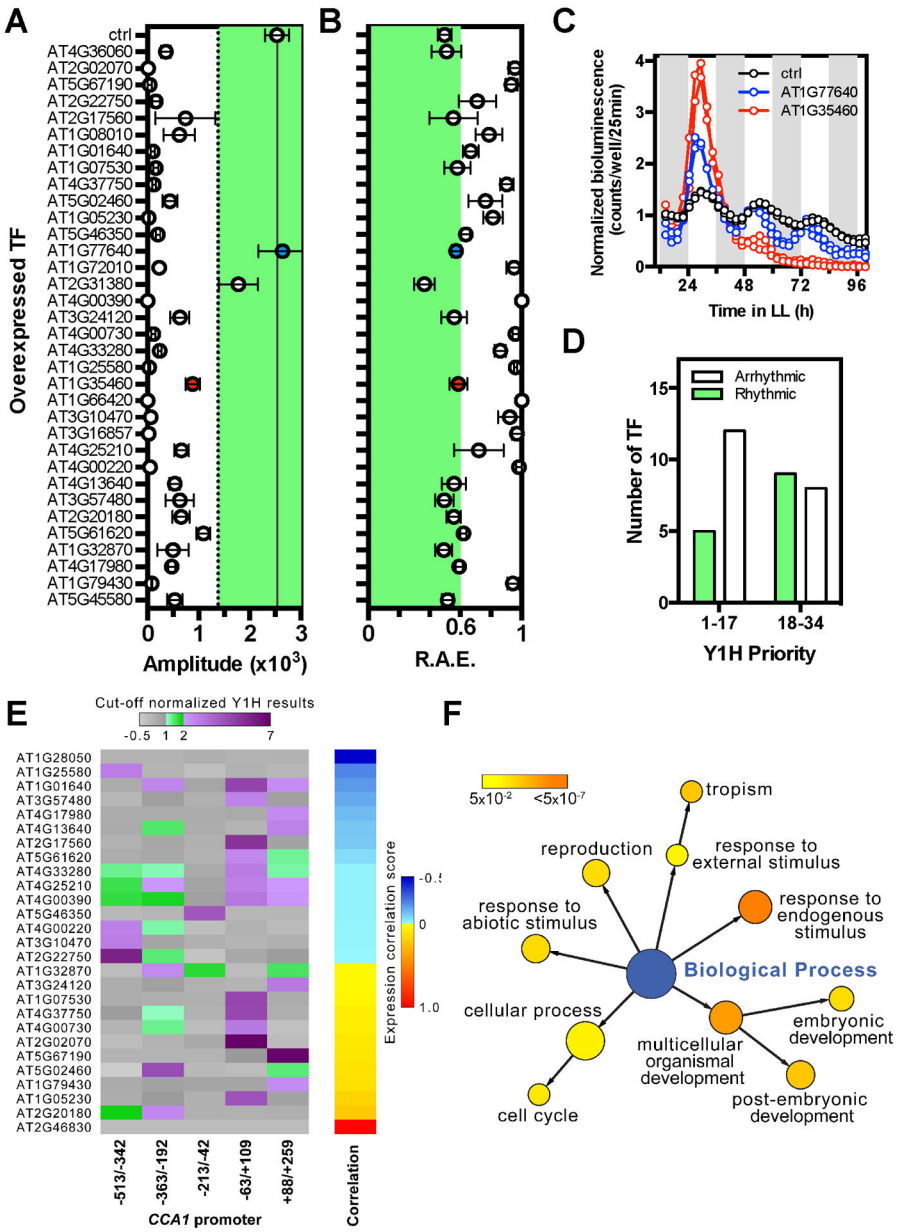

Figure 3. ***In vivo*** **validation of yeast one-hybrid candidates**

A) and **B)** Circadian clock phenotypes for TF overexpression in Arabidopsis protoplasts. TFs are indicated by the corresponding Arabidopsis gene identification numbers shown in (A) and were organized based on β-galactosidase reporter activity (descending order). Each symbol represents the average value ± SEM for amplitude (A) or relative amplitude error (R.A.E.) (B) estimates of *CCA1* promoter driven luciferase expression (*CCA1::LUC+*) in 3 independent experiments (n=3–6). The cut-off limit was set at 3 standard deviations below the average amplitude value for the control (green shaded area indicates the interval of confidence for the control) (A). A relative amplitude error <0.6 (green shaded area) is indicative of rhythmic luciferase activity (B). Colored symbols (red and blue) indicate the examples shown in (C). **C)** Bioluminescence analysis of *CCA1::LUC+* expression in

Arabidopsis protoplasts transformed with *TF* overexpression plasmids (red and blue). Control traces (ctrl) correspond to protoplasts transformed with a control overexpression plasmid (black). Results are representative of three independent experiments. **D)** Frequency of rhythmic and arrhythmic clock phenotypes encountered in two groups of candidates based on the order of priority defined by the β-galactosidase reporter activity. **E)** Expression correlation analysis for *CCA1* and each TF validated in Arabidopsis protoplasts. Heat maps representing the Y1H results defined by the β-galactosidase reporter activity obtained for each screened CCA1 promoter fragment (left), and expression correlation scores (right). TFs are sorted according to their correlation score. **F)** Network of Gene Ontology (GO) biological processes associated to TFs validated in Arabidopsis protoplasts. Node labels correspond to GO slim terms. Node colors denote the degree of correspondence to a GO slim term (enrichment p-value < 0.05) (n=10 TFs).

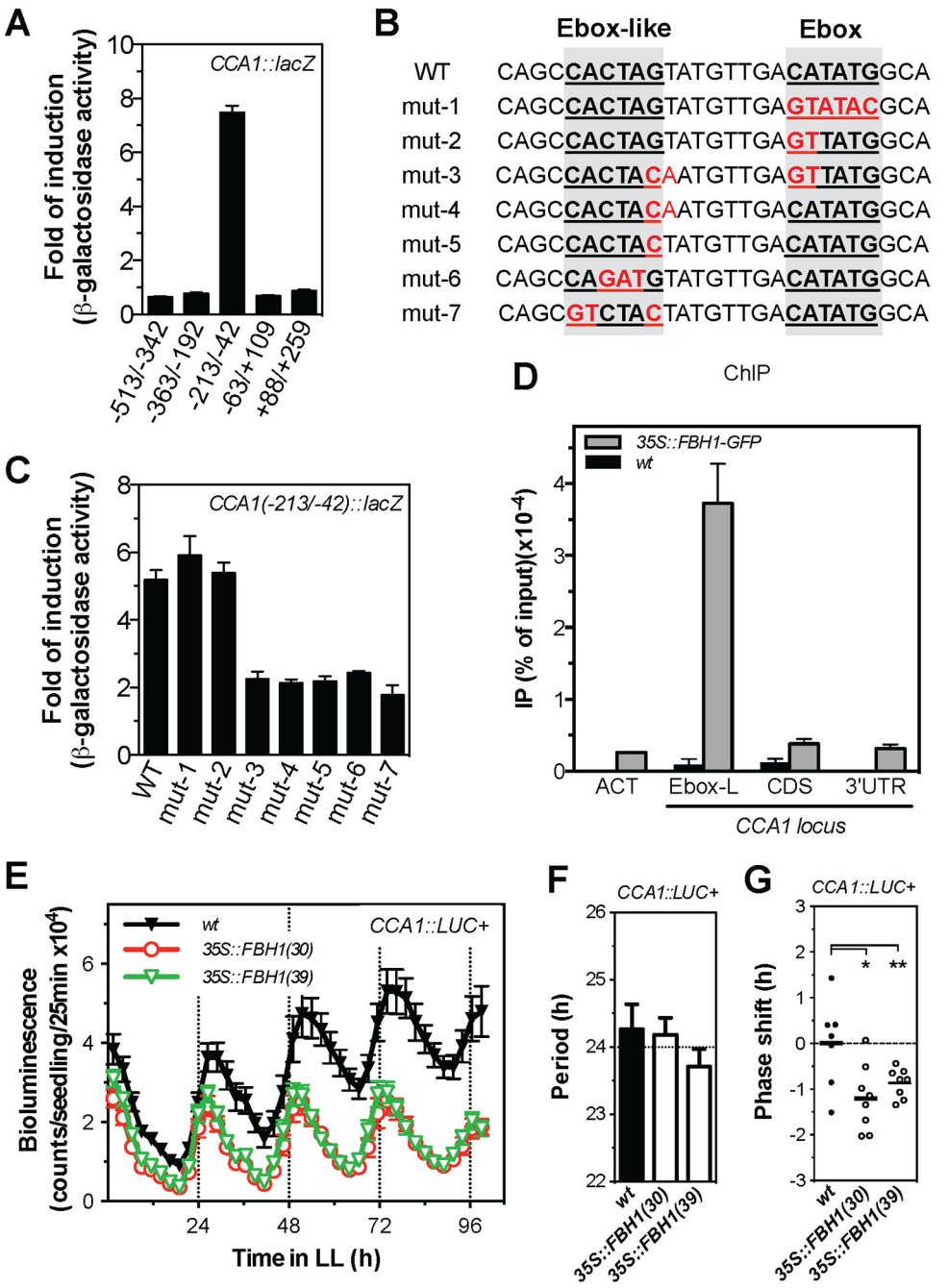

Figure 4. Characterization of the FBH1-*CCA1* promoter interaction

A) Interaction of FBH1 to different regions of the *CCA1* promoter in yeast. Bars represent the fold of induction in β-galactosidase activity for each of the DNA fragments indicated (n=12). **B)** Ebox and Ebox-like motifs present in the *CCA1* promoter region -243/-42. Wild-type (WT) and mutated versions (mut-1 to mut-7) used in Y1H assays (base changes are indicated in red). **C)** Interaction of FBH1 to the - 243/-42 region of the *CCA1* promoter in yeast. Wild-type (WT) and E-box/E-box like mutant (B) promoter fragments fused to the *lacZ* gene were used as promoter baits. Bars represent the fold of induction in β-

galactosidase activity for each of the DNA fragments indicated (n=6). **D)** Binding of FBH1 to the *CCA1* promoter *in vivo*. Chromatin immunoprecipitation assays were performed with *35S::FBH1-GFP* and wild-type *CCA1::LUC+* (*wt*) seedlings. Immunoprecipitated (IP) DNA was quantified by real time quantitative PCR with primers specific for the E-box like locus in *CCA1* promoter (Ebox-L) and for the CCA1 coding (CDS), *CCA1* 3′UTR (3′UTR) and ACTIN (ACT) coding regions as controls. Results were normalized to the input DNA (n=3 independent experiments). **E)** Bioluminescence analysis of *CCA1::LUC+* expression in *FBH1* overexpression lines (*35S::FBH1*) (n=8). Wild-type (*wt*) traces correspond to *CCA1::LUC+* seedlings. Results are representative of at least three independent experiments. **F)** Period estimates of luciferase expression in wild-type *CCA1::LUC+* (*wt*) and the *35S::FBH1* lines shown in (E). Each bar represents the average period value (h) (*wt*, 24.27 ± 0.37; *35S::FBH1(30)*, 24.19 ± 0.25; *35S::FBH1(39)*, 23.71 ± 0.26 [± SD]). **G)** Phase shifts of luciferase expression in wild-type *CCA1::LUC+* (*wt*) and the *35S::FBH1* lines shown in (E). Each symbol represents one seedling and the line is the average phase shift value (*P<0.02, **P<0.03). Values represent means ± SEM (A), (C), (D) or ± SD (D).