Title

The trans-regulatory landscape of gene networks in plants

Journal

Cell Systems, 14(6)

ISSN

2405-4712

Authors

Hummel, Niklas FC

Zhou, Andy

Li, Baohua

et al.

Publication Date

2023-06-01

DOI

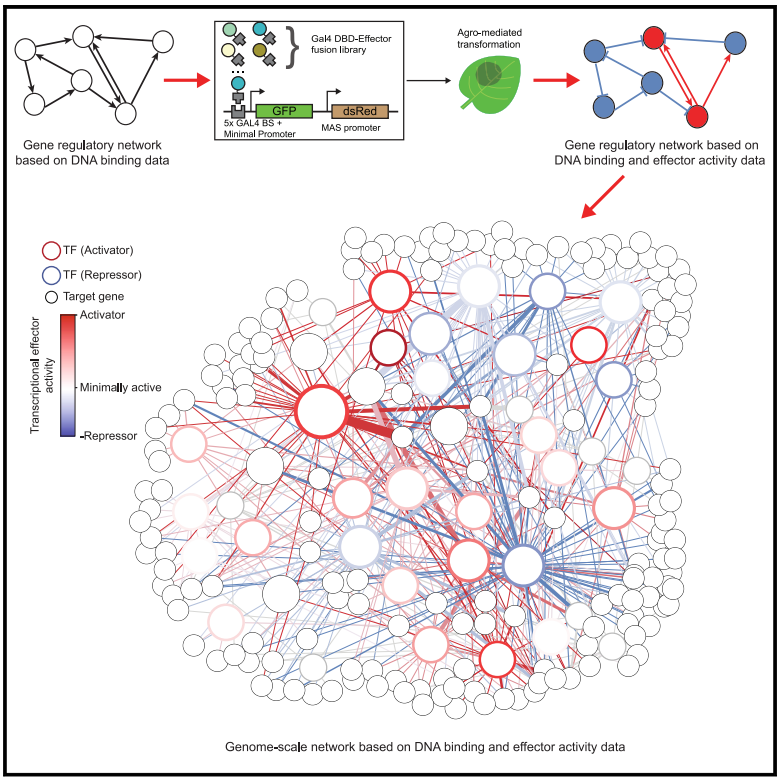
10.1016/j.cels.2023.05.002

Article

Cell Systems

The *trans*-regulatory landscape of gene networks in plants

Graphical abstract



Highlights

- Parallel characterization of hundreds of *Arabidopsis* transcriptional effector domains
- Gene regulatory networks can be enriched with transcriptional effector activity data
- Characterized activators can be utilized for plant metabolic and genome engineering
- Yeast-derived machine learning can correctly localize plant activation domains

Authors

Niklas F.C. Hummel, Andy Zhou, Baohua Li, Kasey Markel, Izaiah J. Ornelas, Patrick M. Shih

Correspondence

pmshih@berkeley.edu

In brief

Hummel et al. utilize a transient plant gene expression platform to characterize the activity of more than 400 transcriptional effector domains from *Arabidopsis*. Using these data, they annotate effector activity in a gene regulatory network based on RNA-seq and TF-DNA binding information to elucidate novel targets for engineering.

Hummel et al., 2023, Cell Systems *14*, 501–511
June 21, 2023
https://doi.org/10.1016/j.cels.2023.05.002

CellPress

Cell Systems

CellPress
OPEN ACCESS

Article

The *trans*-regulatory landscape of gene networks in plants

Niklas F.C. Hummel,[1,2,3,4] Andy Zhou,[1,2,3] Baohua Li,[2,3] Kasey Markel,[1,2,3] Izaiah J. Ornelas,[2,3] and Patrick M. Shih[1,2,3,5,6,*]
[1]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[2]Feedstocks Division, Joint BioEnergy Institute, Emeryville, CA 94608, USA
[3]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94705, USA
[4]Department of Biology, Technische Universität Darmstadt, Darmstadt 64287, Germany
[5]Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA 94720, USA
[6]Lead contact
*Correspondence: pmshih@berkeley.edu
https://doi.org/10.1016/j.cels.2023.05.002

SUMMARY

The transcriptional effector domains of transcription factors play a key role in controlling gene expression; however, their functional nature is poorly understood, hampering our ability to explore this fundamental dimension of gene regulatory networks. To map the *trans*-regulatory landscape in a complex eukaryote, we systematically characterized the putative transcriptional effector domains of over 400 *Arabidopsis thaliana* transcription factors for their capacity to modulate transcription. We demonstrate that transcriptional effector activity can be integrated into gene regulatory networks capable of elucidating the functional dynamics underlying gene expression patterns. We further show how our characterized domains can enhance genome engineering efforts and reveal how plant transcriptional activators share regulatory features conserved across distantly related eukaryotes. Our results provide a framework to systematically characterize the regulatory role of transcription factors at a genome-scale in order to understand the transcriptional wiring of biological systems.

INTRODUCTION

Biological systems are reliant on transcriptional networks, which are largely regulated by transcription factors (TFs). At their core, TFs are defined by two broad functions: (1) specific binding of target *cis*-regulatory DNA sequences through DNA-binding domains (DBDs) and (2) regulating transcription (i.e., gene activation or repression) through their transcriptional effector domains (TEDs). TEDs can serve as biochemical beacons recruiting or inhibiting transcriptional machinery; however, the mechanisms underlying these processes are not well understood and have primarily been studied in other eukaryotes distantly related to plants (i.e., yeast, human, etc.).[1] Recent technical advances and large consortium efforts have dramatically expanded our understanding of TF binding sites across full genomes.[2,3] However, the nature of these interactions has remained elusive, as the functional characterization of TEDs has not been as readily scalable. The regulatory nature of TFs is dictated by their intrinsic protein sequence features, but their tissue- and cell-specific contexts can add another layer of regulation. Furthermore, the physiology of plant cells—especially their cell wall—has made high-throughput studies of libraries of TEDs in plants harder to implement than similar assays in human, fly, or yeast model systems.[4–6] As a result, our knowledge of TEDs that compose the *trans*-regulatory landscape in plants has not kept pace with the characterization of its *cis*-regulatory counterpart.[2,7,8] Hence, identification and characterization of these domains in plants is an important first step toward elucidating the design principles that govern gene regulation in order to ultimately enable more refined approaches to engineer and fine-tune transcription in plants.

Unraveling the functional dynamics of gene regulatory networks (GRNs) is a key challenge of systems biology with the promise to understand the regulatory architecture of biological systems. To observe how genome-scale regulation of transcription occurs, GRNs hinge on the either activating or repressing interactions between individual TFs and their target genes. Hence, a central goal of the field of systems biology is to map genome-scale GRNs to understand the concerted regulation of biological phenomenon and traits.[9,10] However, due to the lack of knowledge of the regulatory role of TFs, GRNs are largely limited to TF binding site and RNA sequencing (RNA-seq)-based co-expression information, which can only indirectly infer whether TF-gene interactions are activating or repressing.[9,11] Therefore, we reasoned that directly measuring the regulatory function of TEDs and integrating this information into GRNs could provide a missing, yet integral, dimension to studying the underlying regulatory wiring of biological systems.

Modulating the expression of plant genes has been a key area of focus for precision crop engineering, as many agronomically

important traits are the result of altered gene expression.[12,13] Hence, the intrinsic *trans*-regulatory elements embedded in plant TF proteins offer a unique resource to mine for TEDs that may advance plant engineering efforts, and understanding their native regulatory role in GRNs could provide targets for engineering. To expand our understanding of plant transcriptional regulation, we systematically measured the activation or repression activity of putative TEDs from over 400 *Arabidopsis thaliana* TFs, providing unique insights into the underlying biochemical properties of plant TEDs. We further show how the integration of *trans*-regulatory information into GRNs can validate and describe the functional role of TFs in gene networks. Our findings demonstrate how genome-wide functional characterization of TEDs can enhance our understanding of the transcriptional regulation of biological systems, both on a biochemical and systems level.

RESULTS

Genome-wide characterization of plant transcriptional effector domains

The *in vitro* DNA-binding activity of 529 *A. thaliana* TFs has been previously reported[2]; however, mapping TF-DNA interactions alone cannot provide information on the regulatory nature of these interactions, limiting our ability to understand key facets of plant gene networks. Previous attempts in large-scale characterization of TEDs in human and yeast models have focused on short-length TEDs ($\leq$80 amino acid).[6,14] While such studies yield short peptides with transcriptional effector activity, they may not fully capture the regulatory activity of the TF. As an example, some activators rely on multiple long subdomains for activity.[15] Thus, in order to provide a more comprehensive understanding of TF function, we instead focused on experimentally characterizing size-unrestrained TEDs of *A. thaliana* TFs whose DNA-binding motifs and downstream targets have previously been mapped.[2,16] For characterization of TEDs, we utilized a transient, synthetic transcriptional system in *Nicotiana benthamiana* that we previously established.[16] While this system cannot resolve tissue-specific activities of TFs, it allows the rapid characterization of large libraries of *trans*-elements and their intrinsic activity in parallel. First, we generated putative TEDs by identifying and excluding conserved DBDs and selectively extracted the longest non-DBD TF protein sequences, which ranged from 27 to 779 amino acids (Figure S1). We then fused these candidate TEDs to the yeast Gal4 DBD, generating a library of synthetic TFs (Table S1). The Gal4 DBD localizes the TED candidate to a plant synthetic reporter composed of a minimal promoter with 5 concatenated Gal4 binding sites driving GFP.[16] By measuring GFP fluorescence in the presence of a synthetic TF and normalizing the signal for basal expression of the reporter by using a constitutively expressed dsRed, we can individually characterize the functional role of TEDs independent of their regular protein context (Figure 1A).

Using this approach we individually characterized 403 synthetic TFs using transient expression in *N. benthamiana* (Figure 1B; Table S2). We identified 166 activator and 53 repressor domains, defining activation and repression as an increase of GFP expression by at least 100% and decrease by at least 40% in comparison to basal expression of the reporter based on statistical thresholds (see STAR Methods). We found 49 activators displaying stronger *trans*-activation activity than the widely used viral activator VP16, with the strongest activator derived from the cold response TF CBF4 increasing GFP expression 16-fold over basal reporter expression (Figure 1B). Our findings demonstrate the potential of transient gene expression in *N. benthamiana* to systematically study TEDs and enable the development of enhanced genetic engineering tools, providing alternatives to broadly used TEDs like VP16.

In order to validate the findings of our assay and support transferability of our data into *A. thaliana*, we compared our observed TED activity with the activity of each parent TF as a repressor, an activator, or both according to previous studies in *A. thaliana*. We found a large overlap between the activity of TEDs in this study and TFs individually studied *in vivo* (Table S3). Of our 166 annotated TEDs with activator activity, 90 have been previously reported to directly activate expression, and only 7 act as repressors. Of our 52 TEDs with repressor activity, 21 have been previously shown to act as repressors and 2 as activators. In our validation dataset (Table S3), there are two TFs that have been characterized as both activators and repressors, WUSCHEL and bHLH104. In our assay, we characterized WUSCHEL as a repressor and bHLH104 as an activator, which may be a result of both the *cis*- and *trans*-regulatory context of our synthetic system. Annotated activators and repressors with activities conflicting with observations in the literature are likely caused by similar context dependencies. Nonetheless, most TEDs that had previous data on their regulatory activity seemed consistent, suggesting that the intrinsic protein features and regulatory roles of TEDs can largely be reproduced in our system. Notably, we discovered 68 activators and 28 repressors, highlighting how our approach enables the discovery and characterization of TEDs. The broad overlap of referenced TF activity with the TEDs described here indicates the consistency and transferability of TED activity between our synthetic heterologous system and *in vivo* observations in the native plant. To further validate our repressor findings, we studied the occurrence of the well-known ethylene-responsive element binding factor-associated amphiphilic repression (EAR) motif in our TED populations. We found an overrepresentation of the repressive EAR motif in the repressor population (33% of TEDs with motif) when compared both to the activator (4%) and minimally active populations (13%, Figure S2), supporting the agreement between our assay and the *in vivo* function of TEDs in their natural context. Taken together, we found a large qualitative agreement between our TED characterization in *N. benthamiana* and published *in vivo* activity in *A. thaliana*, which enabled us to combine DNA-binding and TED activity for further analysis.

Our dataset spans TEDs from 34 TF families and allows us to study functional trends across TF families (Figure 1C). For example, the AP2-EREBP TF family comprises 147 TFs in *A. thaliana*, and of the 70 AP2-EREBP TFs studied here, only 4 TEDs act as repressors and 54 TEDs as activators (Figure S3). This indicates a bias toward activators inside the AP2-EREBP family. Conversely, in the C2H2 TF family we found 8 out of 20 TEDs studied here to significantly repress gene expression with none as characterized activators. These observations overlap with human C2H2-TFs, which mostly contain repressive TEDs.[17] Thus, regulatory roles across long evolutionary

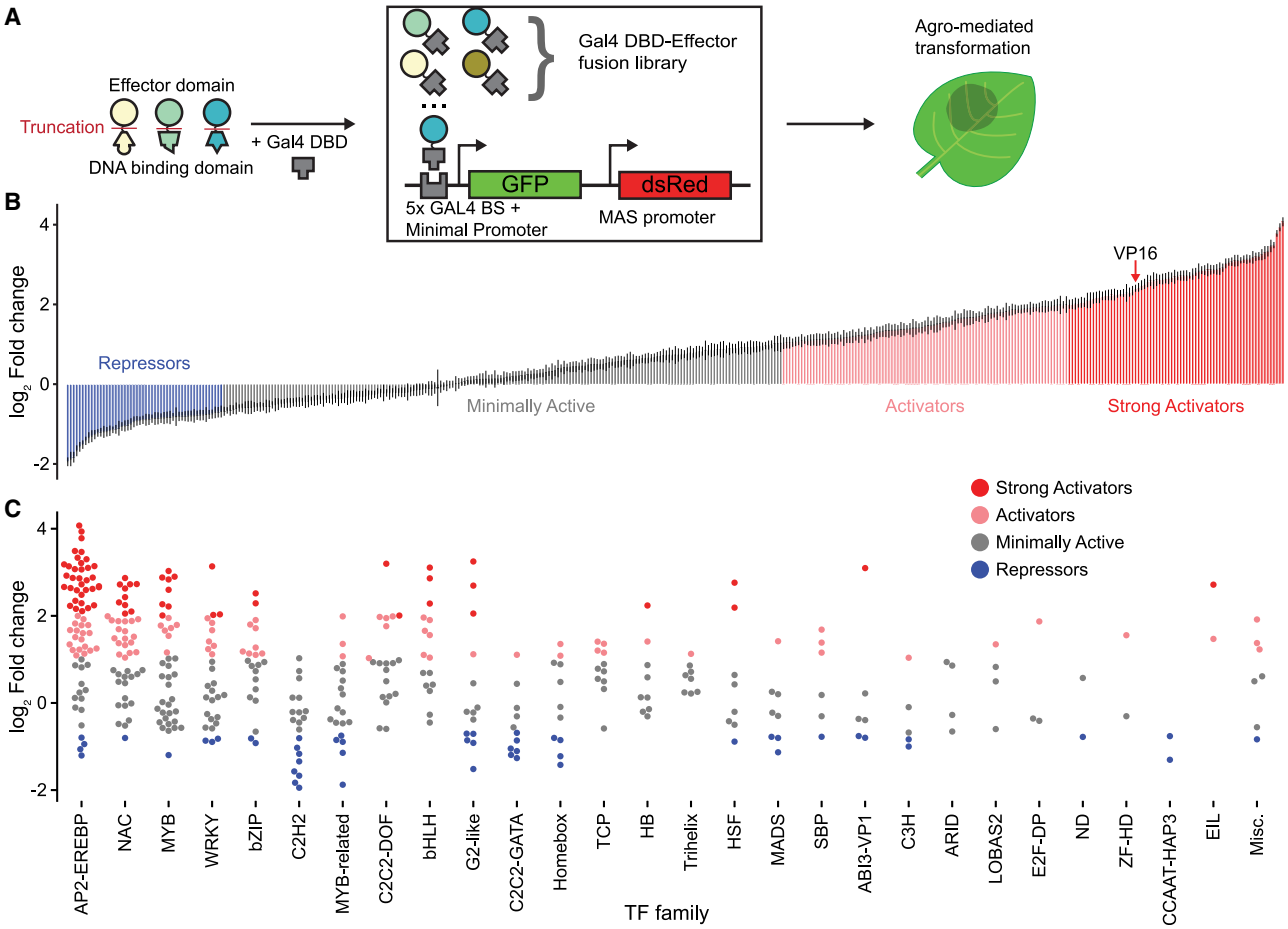


Figure 1. Genome-scale characterization of hundreds of plant transcriptional effector domains

(A) Truncated putative TEDs are fused to the yeast Gal4-DBD to generate a library of synthetic TFs. Gal4-TED fusions bind to a synthetic promoter, and their effect on transcription is measured via a fluorescent reporter GFP and normalized using a constitutively expressed dsRed.

(B) Normalized GFP expression of 403 synthetic TFs in relation to background reporter expression in *N. benthamiana* leaves 3 days post infiltration (n = 16 biological replicates). Arrow indicates the position of Gal4-VP16 as a strong activator control.

(C) Normalized GFP expression from (B) grouped by TF family. Individual data points represent single TEDs characterized in this study. TF families with single TEDs in the screen were grouped into Misc. and include REM, LIM, RAV, zf-GRF, RWP-RK, BSD, and NLP.

distances might be conserved in the C2H2 family. TF binding sites within the same family are often redundant and general trends of specific TF families as either activators or repressors may increase the robustness of the regulatory network through a form of functional redundancy.

Trends in TF downstream targets based on TED activity

Combining TF binding site and TED activity information allowed us to study trends in which TF-gene interactions are either repressive or activating. By incorporating this regulatory logic, it becomes possible to describe recurring network motifs that are broadly found across biological systems.[18] One such network motif in prokaryotes is negative autoregulation (NAR), where a repressor downregulates its own expression.[19] NAR enables the acceleration of response times and reduces cell-to-cell variation in protein concentration, thus enabling robust regulation of their targets.[18,20] To investigate usage of NAR in plant TFs, we combined TED activity with published DNA-binding data.[2] When comparing TED populations based on their regulatory activity, we found that repressors were more likely to autoregulate than strong activators (defined as TEDs increasing gene expression more than 400%, Fisher's exact test p = 0.04, Figure S4A). We also searched for a bias between positive and negative feedback loops, i.e., two TFs regulating each other, but did not observe any bias between the activator and repressor populations (Figure S4B). Our analysis of NAR in plants supports convergence of these motifs across both prokaryotes and eukaryotes, supporting previously suggested emergent properties of transcriptional regulation.

We further identified trends in the functional roles of genes targeted by TFs with strongly activating TEDs. We analyzed the gene ontology (GO) terms of genes targeted by activators stronger than VP16. We found that the GO terms of these genes were enriched for terms linked to response to hormones, stresses, and external stimuli. GO terms linked to primary or secondary metabolism were depleted, except for terms linked to cell-wall

biogenesis (Figure S5). This suggests a requirement for strong gene activation to enact the rapid changes to transcriptional programming needed for a concerted response to stimuli rather than the direct activation of metabolic pathway genes during housekeeping functions. To thoroughly explore this concept, future studies of genetic perturbations of these activators in *A. thaliana* will be necessary to study their role in the response to external stimuli in detail.

An integrated *cis*- and *trans*-regulatory gene network elucidates the functional dynamics of TF-gene interactions

GRNs integrating *cis*-regulatory information and RNA-seq attempt to elucidate the underlying regulatory logic of TFs and their targets[11,21]; however, these GRNs are limited to indirect inferences of TF activity, rather than directly assaying how target TFs modulate gene expression. Thus, integrating *trans*-regulatory TED data as a layer on top of established GRNs could enhance their explanatory power and help study how TF-DNA interactions translate into regulatory output. As an ideal case study for this approach, we chose to investigate the well-characterized transcriptional response to nitrogen in *A. thaliana.*[9]

The uptake and sensing of nitrogen in plants relies on an intricate network of TFs and the modulation of single TFs' regulatory activity can have stark growth impacts, highlighting the potential for engineering nitrogen signaling in plants.[22] Recently, Varala et al. generated a high-confidence nitrogen-responsive GRN encompassing 37 TFs and 171 direct genomic targets by combining published TF-DNA binding with temporal RNA-seq.[11] Using our TED data, we annotated the links between TFs and their targets in this network as activating or repressing, thereby generating the first GRN integrating TED activity with DNA binding and temporal RNA-seq data (Figure 2A; Table S4).

Annotating TF-gene interactions with regulatory activity relies on the assumption that (1) the transcriptional effect of our Gal4-TED fusions on the reporter construct resembles the endogenous effect of its parent TF on its targets and (2) the TF-DNA binding used for establishing the GRN reflects actual interactions of a TF with the respective target promoter. We therefore sought to verify the reliability of both the transfer from synthetic TF activity to native TF and the precision of the underlying DNA binding by testing whether the activity of full-length TFs interacting with native promoter elements is consistent with our measured TED activities in a synthetic transcriptional system. We reconstituted native TF-DNA interactions of core nitrogen metabolism genes by building GFP reporter constructs driven by the *Arabidopsis* promoters of nitrate reductase 1 (NR1) and nitrite reductase 1 (NIR1) (Figure 2B). We then co-expressed full-length TFs that are predicted to regulate these genes throughout the entire time course according to the GRN. We found that the TFs CRF10 and AT1G12630—both annotated as activators in our assay and predicted to interact with NIR1—significantly increased the activity of the NIR1 promoter (Figure 2C). Five TFs whose TEDs are activators in our assay—CRF4, bZIP3, TGA4, ANAC018, and AT1G12630—are predicted to interact with the NR1 promoter in the GRN. All five TFs altered expression with AT1G12630 strongly inducing; CRF4, bZIP3 and ANAC018 weakly inducing; and TGA4 repressing the NR1 promoter (Figure 2D). CRF4, bZIP3, and TGA4 have been previously shown to induce gene expression during nitrogen response.[23] Hence, we were able to validate the regulatory activity of 6 out of 7 TF interactions with core nitrogen promoters. The repression of transcription by TGA4 may be a result of lacking co-activators, or secondary mechanisms that are found in the native system, showing the limitations of our transient expression system compared to testing the TFs in its native context in *A. thaliana*. Nevertheless, the fact that our heterologous system can elucidate activity by observing single TF-promoter interactions confirms that GRNs can be enriched by annotating TED activity.

To further validate that TFs in GRNs can be annotated as activators or repressors based on TED activity, we sought to exploit the temporal dynamics of target genes in the nitrate response network by Varala et al.[11] For example, genes targeted by TFs whose TED is an activator in our assay should tend to increase in mRNA abundance over time. By leveraging the temporal nature of the Varala et al. dataset, we validated our measured TED activities and observed the causal relation between TF and the transcriptional output of their downstream targets. Nitrate-responsive genes show altered gene expression early after nitrate induction.[24] Therefore, we focused on the early nitrogen response between 0 and 30 min. At 15 min post nitrate induction, we observed a set of six annotated activators that target primary nitrate response genes (NR1, nitrate reductase 2, and NIR1) (Figure S6). The presence of activators should lead to induction of target genes. NR1/2 and NIR1 indeed show increased levels of expression at later time points (Figure S6B) with increased expression after the interaction with our characterized activators including the validated bZIP3 and AT1G12630. We further observed the RNA abundance of all genes targeted by this activator group (Figure S6C). As a measure for dynamic changes in gene expression we calculated the rate of expression change in between every time point for every gene of this group (Figure S6D). We found that between 20 and 30 min the majority of genes targeted by the activators present in the 15 min sub network showed their largest rate of expression induction (Figure S6D), and none showed their strongest reduction of expression (Figure S6E). This indicates that nitrate-responsive genes targeted by this activator group show the predicted response based on *trans*-regulatory activity *in vivo.*

Together our results demonstrate how integrating TED activity from heterologous experiments into GRNs can help recapitulate the regulatory relationships of TFs with their downstream targets. To verify the systemic effects of the described activators, future studies that can assess their regulatory activity within a native host and tissue-specific manner may provide more insight into the nuances of their regulatory role; however, the technology that enables this at a high-throughput scale will first need to be developed. Overall, our characterization of TEDs provides an important first step toward filling major gaps in our knowledge of GRNs that top-down observations have been unable to resolve.

Plant activators expand modularity of synthetic biology and genetic engineering tools

Having shown that TED activity can enrich GRNs, we applied our TEDs in a synthetic biology context to control gene expression and expand the dynamic range of native gene transcriptional profiles. Previously developed plant synthetic biology tools have heavily relied on a small subset of characterized effectors.

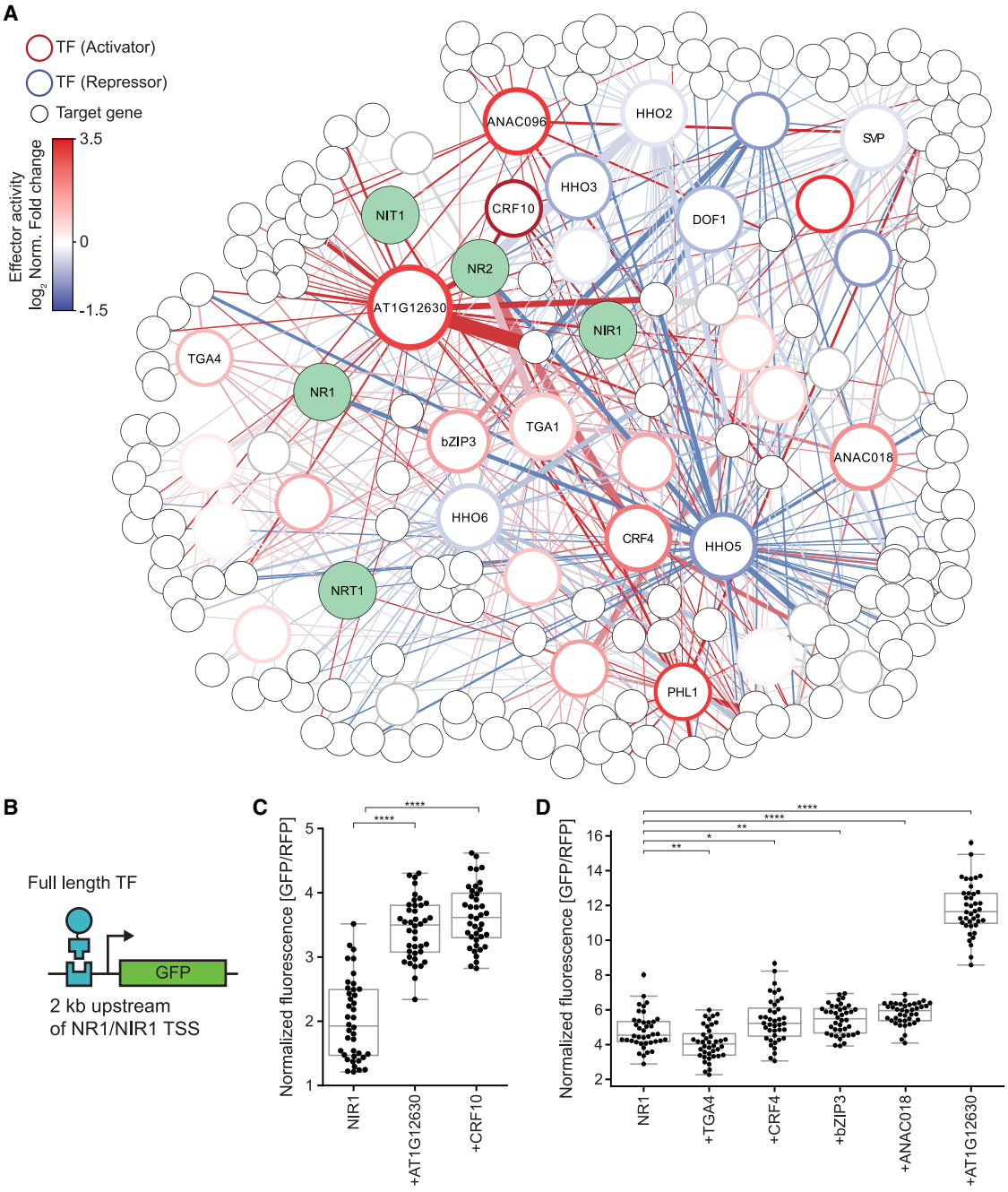


Figure 2. An integrated *cis*- and *trans*-regulatory GRN on nitrogen response in *Arabidopsis*

(A) GRN describing both TFs and respective target genes showing significant temporal changes in RNA abundance in response to nitrate in *A. thaliana*. Hollow nodes with colored edges depict TFs with respective *trans*-regulatory activity from this study. Target genes are small white nodes with black borders. Edges are based on experimentally verified DNA binding of TFs to target genes and annotated with experimental TED activity data (edge color) and the predicted influence of a TF to its target (edge width)[11]. Green nodes indicate core nitrogen metabolism genes.

(B) Coexpression of native full-length TFs enables modulation of GFP expression driven by native NR1/NIR1 promoter regions. TSS, transcriptional start site.

(C and D) Normalized GFP fluorescence for coexpressed full-length activators with promoters derived from (C) NIR1 and (D) NR1 for n = 40 biological replicates. Asterisks indicate Mann-Whitney U test *$p \leq 5 \times 10^{-2}$, **$p \leq 5 \times 10^{-3}$, ****$p \leq 5 \times 10^{-5}$.

For example, the herpes simplex virus-based VP16 domain has been utilized in many eukaryotic systems, including plants, as the state-of-the-art activator since its discovery over 30 years ago.[25,26] Thus, it is of note that many of our characterized TEDs demonstrate stronger activation activity than the VP16 domain, which is commonly used in genome engineering approaches (e.g., dCas9-based CRISPR activation, synthetic TFs, etc.).[27–29] Our findings demonstrate how TED screens

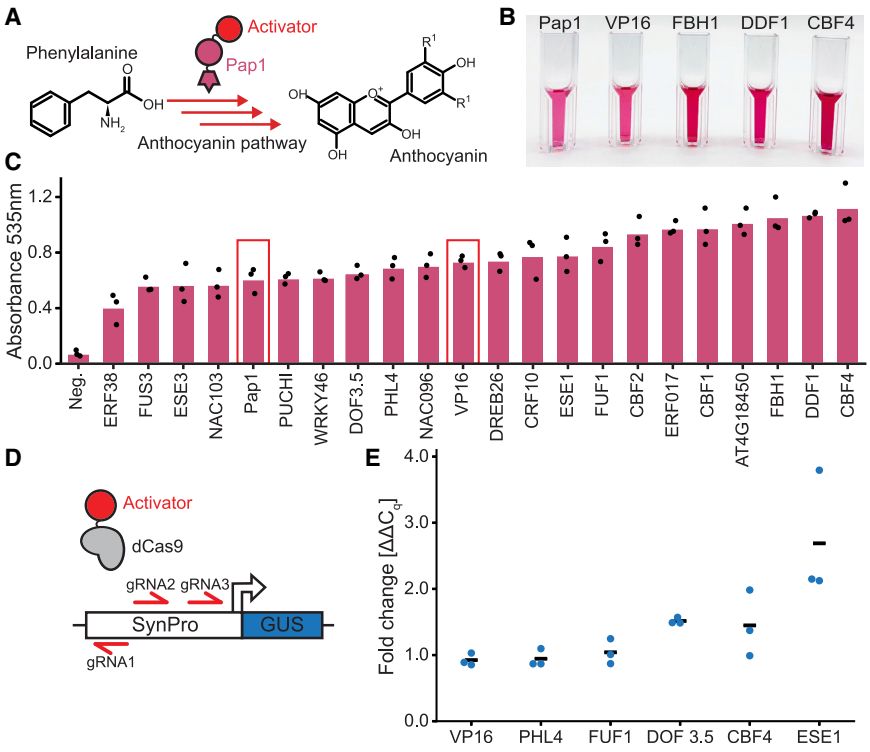


Figure 3. Strong plant activators outperform VP16 when integrated into synthetic biology applications

(A) Fusion of strong activators to the anthocyanin master regulator PAP1 promotes production of anthocyanins.

(B) Visual representation of anthocyanin extracts quantified in (C).

(C) Quantification of anthocyanins extracted from *N. benthamiana* leaf tissue expressing PAP1-fusion constructs (n = 3 biological replicates). Red bars indicate Pap1 as a basal expression and Pap1-VP16 as a positive control. Neg. : Leaves infiltrated with strain GV3101 with no binary vector as negative control.

(D) Activator fusion to dCas9 to modulate target gene expression. Three gRNAs localize the dCas9-activator fusion to a synthetic promoter driving GUS. (E) Fold change in transcript abundance of dCas9-activator fusions relative to the reporter construct alone quantified by qRT-PCR (n = 3 biological replicates).

enable a powerful approach to mine for host-specific (e.g., plant-specific) activator domains that can be superior to the state-of-the-art domains currently utilized.

To explore the utility of our discovered activation domains for metabolic and genome engineering, we tested how 20 activator domains stronger than VP16 would perform when fused to other TFs as a means to enhance their transcriptional output. We fused our chosen activator domains to the anthocyanin master regulator PAP1 as it activates the expression of multiple anthocyanin pathway genes resulting in a quantitative readout via elevated levels of anthocyanins in plant tissue (Figure 3A).[30,31] We expressed PAP1-activator fusions in *N. benthamiana* for 3 days and quantified the anthocyanin content by absorbance measurements of extracts from leaf tissue. Multiple activators showed increased concentration of anthocyanins in comparison to PAP1 and a PAP1-VP16 fusion (Figures 3B and 3C). Eight PAP1-activator fusions showed significantly higher absorbance values than PAP1 and seven significantly higher than PAP1-VP16 (two-sided t test, $p < 0.05$), which indicates a trend toward enhanced metabolic output of anthocyanins by PAP1-activator fusions (Table S5). The strongest activator TED from the screen, derived from CBF4, achieved the strongest increase in anthocyanin concentration. Taken together, our panel of top activation domains can be easily integrated into synthetic biology applications like the optimization of the transcriptional output of TF master regulators.

The modularity afforded by deactivated RNA-guided nuclease variants (e.g., dCas9) allows for the targeted alteration of gene expression when selectively defined by engineered guide RNAs.[28,32] Thus, the versatility of the DNA-binding capability of dCas9-activator constructs has been leveraged to enable genome wide CRISPR activation screens, but again have mostly relied on VP16-based viral activators. Hence we sought to benchmark our top activator candidates against VP16 in an established expression system.[33] We fused five strong activators from our screen to dCas9 and compared these dCas9-activator fusions to dCas9-VP16 by targeting them to a synthetic promoter (Figure 3D). We quantified transcript abundance by qRT-PCR with RNA extracted from *N. benthamiana* leaf tissue 3 days post Agro-mediated transformation. We observed that dCas9-VP16 displayed extremely low activity in comparison to two activator domains from ERF38 and DOF3.5, with ERF38 displaying significant increase of expression (Dunn's test $p < 0.005$; Figure 3E; Table S6); thus, our characterized domains have the potential to enhance CRISPR activation in plants. The field of genome engineering has embraced the use of VP16-based activators and has largely coped with its low activation activity by recruiting large numbers of VP16 via various strategies.[34,35] As an alternative, our TED characterization demonstrates how host-derived TEDs can result in an increased dynamic range of gene expression and expand the available TEDs that can be utilized for modifying transcription. Ultimately, our genome-wide screen enabled us to identify strong activator domains that can be used to tunably enhance transcription in a genome-specific manner, thereby providing a foundation for rapid generation of functional genomics toolsets.

Conserved activity of plant transcriptional activators across eukaryotes

Just as the function of VP16 can cross eukaryotic super families,[33,36] plant transcriptional activators can utilize molecular machinery and mechanisms broadly conserved between distantly related species, which demonstrates how plant activators can be implemented as orthogonal tools in other eukaryotes. In order to investigate the conservation of the regulatory activity of plant activators into other eukaryotes, we tested the ability of 20 activators stronger than VP16 to promote constitutive gene expression in the

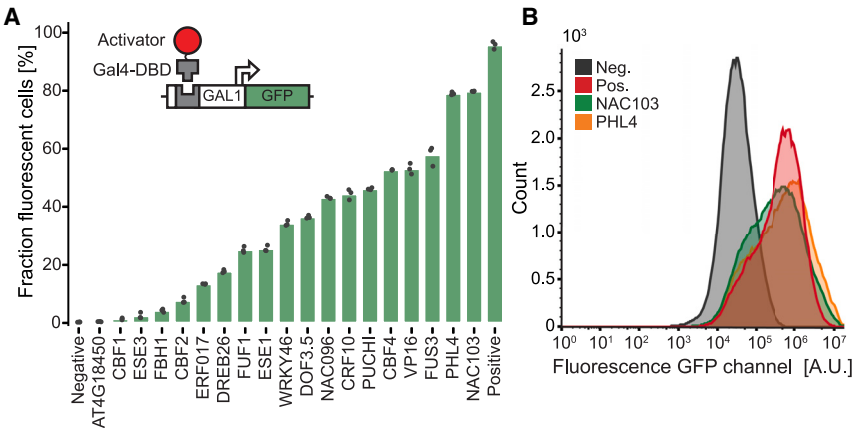


Figure 4. Plant activator activity is conserved in yeast

(A) Plant activators can induce a native yeast promoter when fused to the GAL4-DBD. Fractions of cells showing fluorescence in the repressed state of the GAL1 promoter grown in glucose. Each datapoint represents above fluorescence threshold frequency for 100,000 recorded events.

(B) GFP fluorescence intensity distributions of activator and control populations for 100,000 recorded events.

model yeast system, *Saccharomyces cerevisiae*. We designed an expression cassette utilizing the well-characterized yeast-inducible GAL1 promoter, which is induced in presence of galactose, repressed by glucose, and contains Gal4-binding sites,[37] driving the fluorescent reporter GFP. We then observed the ability of Gal4-DBD-TED fusions to induce gene expression in the repressed state of the promoter using flow cytometry (Figure 4A). TED activity was quantified by measuring the fractions of cells whose GFP expression was equal to or higher than that of GAL1-GFP induced by galactose, while excluding observations similar to GAL1-GFP in glucose. When the Gal4-DBD-TED fusions were expressed constitutively, GFP expression was observed in <1%–80% of the cell populations (Figure 4A; Table S7). The TEDs derived from NAC103 and PHL4 were able to outperform VP16, marking them for further optimization in fungi (Figure 4B). The Gal4-DBD-activator fusions were tested in the presence of glucose, the repressed state of the GAL1 promoter. Still, multiple activators were able to enhance GFP expression, highlighting their potential for developing activation tools. We further observed that although some TF families like the AP2-EREBP TF family are only found in plants,[38] activators from this family function in yeast, suggesting that, while evolved uniquely in plants, disparate TF families may have converged on similar mechanisms of activation.

The observation that our plant activators function in fungi suggested that plant activators broadly utilize the same conserved general eukaryotic mechanisms underlying transcriptional activation. We therefore analyzed the protein sequences of our TEDs using ADpred, a machine learning model trained on a large set of putative activation domains in 30-amino-acid-long protein sequences in *S. cerevisiae*.[39] We calculated the ADpred score for 30 amino segments of all TEDs in this study as described and assigned a binary value to every TED depending on whether it contained an amino acid section with an ADpred score $\geq$ 0.9. We found that activators are more likely to contain consecutive amino acid residues predicted to be activation domains than the repressor and minimally active populations (Figure 5A; two-sided Fisher's exact test, p = 0.00012). We found motifs in 19 out of our 20 chosen activators stronger than VP16 that scored above our ADpred threshold (Figure S7). We extracted ADpred-predicted subsections of three TED regions with strong activator activity (Figure 5B) and benchmarked them against their full-length TEDs and VP16 in *N. benthamiana* using the same assay as described in Figure 1A. The ADpred-predicted motifs of ESE3 and WRKY46 induced the expression of GFP similar to their full-length TEDs and outperformed VP16 (Figure 5C), showcasing the potential to mine plant TFs using the ADPred model. The two motifs of PHL4 were not able to induce GFP in the same manner as their parent TED, suggesting that either the two motifs need to function as a bipartite motif or the parent TED uses a mechanism that the model cannot predict.

Because we demonstrate that machine learning models trained on yeast activation domains could reliably predict plant activation domains, we investigated whether we could observe similarities in biochemical features between plant and yeast activators. Specifically, in yeast there have been well-documented biases for acidic and large hydrophobic amino acid residues found in activation domains.[40] We indeed found strong biases in the amino acid composition between the repressor and activator populations (Figures 5D and S8A). Acidic, hydrophobic, and aromatic residues were significantly overrepresented among our characterized activators, whereas basic residues (e.g., arginine, lysine, and histidine) were significantly depleted and only enriched in repressors. These biases match the sequence profile of acidic activation domains found in mammalian and yeast systems.[1,39,41,42] The isoelectric point of TED populations also differed, supporting the importance of the amino acid composition. Activators show low isoelectric points, whereas repressors exhibit a wide range of isoelectric points. These data suggest that overall charge is a more important feature of activators than of repressors in plants (Figure 5E). Structural disorder has also been suggested to be linked to TED activity, but while TEDs were predicted to be on average > 75% disordered, we did not observe a bias between TED populations (Figure S8B). We further couldn't find a bias between the protein length when comparing TED populations (Figure S8C). Taken together, our results indicate that plant activators share sequence features with their counterparts from distant eukaryotes, suggesting the utilization of a general eukaryotic mechanism for transcriptional activation.

DISCUSSION

A holistic understanding of GRNs necessitates both *cis*- and *trans*-regulatory information, yet the vast majority of studies have largely focused on TF-DNA interactions, while omitting the regulatory nature of these interactions. As a result, this has precluded our ability to reconstruct the true regulatory architecture and transcriptional logic underlying gene networks. To

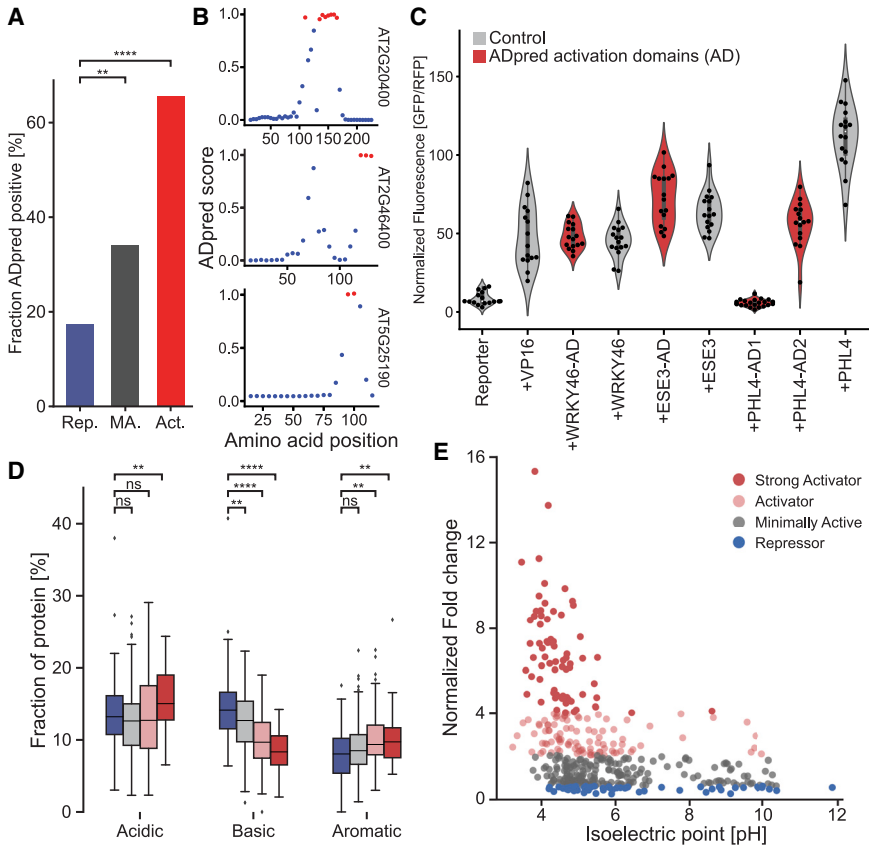


Figure 5. Translation of a yeast machine learning algorithm on plant TEDs supports trends in conserved biochemical features of activators

(A) Plant activators are enriched in activation domains predicted by the yeast machine learning model ADpred. Sum of TEDs with ADpred score $\geq$ 0.9 of each TED population. Asterisks indicate Fisher's exact test **p $\leq$ 5 × 10^{-3},****p $\leq$ 5 × 10^{-5}. Rep., repressor; MA., minimally active; Act., activator.

(B) ADpred analysis of three strong plant activators. ADpred scores were calculated for every 30-amino-acid stretch slid along the protein sequence with window size = 5. Red dots indicate 30-amino-acid-long segments with ADpred score $\geq$ 0.9, blue dots < 0.9.

(C) ADpred predicted activator motifs can perform similar to full-length TEDs. Distribution of normalized GFP fluorescence reads for n = 16 biological replicates in *N. benthamiana*. Motifs found using ADpred indicated in red. M, motif.

(D) Amino acid frequencies of all individual candidate TEDs grouped into their respective population (asterisks indicate Mann-Whitney U significance test *p $\leq$ 5 × 10^{-2}; **p $\leq$ 5 × 10^{-3}; ***p $\leq$ 5 × 10^{-4}; ****p $\leq$ 5 × 10^{-5}; ns, non significant).

(E) Isoelectric point of TEDs mapped to performance in TED screen.

address this fundamental gap in our knowledge, we directly and empirically measure the regulatory role of hundreds of *Arabidopsis* TFs, enabling the first integrated genome-scale *cis*- and *trans*-regulatory analysis to infer gene network behavior. By layering on *trans*-regulatory activities of TFs, we describe an approach to enrich our systems-level understanding of transcriptional regulation in any biological systems by providing directionality and causality in transcriptional networks.

Although our system characterizes TEDs in a non-native context, many of our described regulatory activities are consistent with what has previously been characterized in prior studies (Table S3). Nonetheless, TFs have shown myriad ways of regulating transcription through post-translational modifications (e.g., phosphorylation), heterodimerization, or the interaction with a complex set of coactivators and corepressors.[43,44] These extra layers of regulation can lead to conflicting activities of TFs in different tissues, where a given TF may act as an activator in one tissue and as a repressor in another. As a prominent example, WUSCHEL, the TF regulating stem cell maintenance in shoot and floral meristems, shows activity as either a repressor or activator depending on the tissue it is expressed in, highlighting the importance of tissue-specific observations.[45] From this premise, it complicates the establishment of high-throughput characterization efforts of TEDs and their integration of this knowledge into transcriptional networks. Nonetheless, we find that the majority of our findings are validated by previous studies and *in vivo* observations (Table S3), supporting the utility of our approach.

When generating our library of synthetic TFs, we extracted TEDs that are positioned both N and C terminally in their endogenous TF but studied them as C-terminal fusions in our synthetic TF system. While we show that N-terminal TEDs indeed show activity as C-terminal fusions, we cannot fully rule out that there are positional effects on TED activity. In our assay, we utilized a minimal synthetic WUSCHEL promoter to establish basal expression that can be modulated by our TF library. The *cis*-element itself can have an effect on the potential activation or repression that can be achieved. Thus, while we cannot rule out that the strength of our TFs might vary with a different promoter, our results with the native NR1 and NIR1 promoters demonstrate that overall activity of activators can be conserved when assayed in our system. Future work focused on systematically scaling this approach to study all TFs across entire genomes and iterations on context dependency will provide invaluable information to enrich our understanding of GRNs and transcriptional regulation.

Our findings add a new perspective on the complex evolution of TFs and the emergence of transcriptional wiring in biological systems. The functionality of plant TEDs in yeast suggests that core mechanisms for transcriptional activators are deeply conserved across eukaryotes. This trend is maintained even for TF families that have uniquely evolved in plants (i.e., the AP2-EREBP family), revealing how distinct TFs have converged upon shared biochemical features within the boundaries of universal transcriptional mechanisms. More broadly, this suggests a model where DBD of TF families may evolve somewhat

independently from their TEDs, enabling extreme changes in regulatory activity in even closely related TFs. Overall, our study provides a new perspective on how to investigate the evolution of TF function, which will help reveal how complex phenotypes may have evolved.

Many important agricultural traits are dictated by TFs, thus presenting key targets for selectively modulating and engineering solutions to challenges in agriculture, bioenergy, and sustainability. Our findings provide the foundational knowledge needed to systematically map the regulatory role of TFs at a genome scale in order to elucidate the underlying genetic wiring of plant transcriptional networks. This systems-level understanding of plants will be an invaluable resource to the broader plant biology community in understanding the circuitry and dynamics that control nearly all facets of plant physiology, development, and responses to the environment.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - *N. benthamiana* growth conditions
  - Bacterial and yeast growth conditions
- METHOD DETAILS
  - Design of TED candidates
  - Construct design and assembly
  - Agro-mediated transient transformation of *N. benthamiana*
  - Quantification of anthocyanin content
  - Quantitative reverse transcription PCR experiments
  - Flow cytometry
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Grouping of TED populations
  - Gene ontology enrichment of activators and autoregulation
  - Generating a *trans*-regulatory annotated nitrogen response GRN
  - Validation of GRN predicted TF-Gene interactions
  - Localization of activation domains in plant TFs using ADpred

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cels.2023.05.002.

ACKNOWLEDGMENTS

We thank Ronan O'Malley for providing template DNA for the amplification of all TED regions of interest and both Simon Alamos and James Nunez for reviewing the manuscript. This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

AUTHOR CONTRIBUTIONS

Project design: N.F.C.H., B.L., and P.M.S.; experimental work: N.F.C.H., A.Z., B.L., K.M., and I.J.O.; analyses: N.F.C.H. and A.Z.; visualization: N.F.C.H. and A.Z.; financial support: P.M.S.; supervision: P.M.S.; writing – original draft: N.F.C.H., A.Z., and P.M.S.; writing – review & editing: N.F.C.H., A.Z., K.M., and P.M.S.

DECLARATION OF INTERESTS

P.M.S. and N.F.C.H. have a patent pending for the library of synthetic TFs and all derived parts under US patent application number 18/298,942.

Received: November 1, 2022
Revised: March 21, 2023
Accepted: May 11, 2023
Published: June 21, 2023

REFERENCES

1. Brzovic, P.S., Heikaus, C.C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R.E., and Hahn, S. (2011). The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. Mol. Cell *44*, 942–953.
2. O'Malley, R.C., Huang, S.-S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. Cell *165*, 1280–1292.
3. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
4. Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., and Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature *528*, 147–151.
5. Sanborn, A.L., Yeh, B.T., Feigerle, J.T., Hao, C.V., Townshend, R.J., Lieberman Aiden, E., Dror, R.O., and Kornberg, R.D. (2021). Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. Elife *10*, e68068.
6. Tycko, J., DelRosso, N., Hess, G.T., Aradhana, Banerjee, A., Mukund, A., Van, M.V., Ego, B.K., Yao, D., Spees, K., et al. (2020). High-Throughput Discovery and Characterization of Human Transcriptional Effectors. Cell *183*, 2020–2035.e16.
7. Marand, A.P., Chen, Z., Gallavotti, A., and Schmitz, R.J. (2021). A cis-regulatory atlas in maize at single-cell resolution. Cell *184*, 3041–3055.e21.
8. Jores, T., Tonnies, J., Wrightsman, T., Buckler, E.S., Cuperus, J.T., Fields, S., and Queitsch, C. (2021). Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. Nat. Plants *7*, 842–855.
9. Gaudinier, A., Rodriguez-Medina, J., Zhang, L., Olson, A., Liseron-Monfils, C., Bågman, A.M., Foret, J., Abbitt, S., Tang, M., Li, B., et al. (2018). Transcriptional regulation of nitrogen-associated metabolism and growth. Nature *563*, 259–264.
10. Chen, D., Yan, W., Fu, L.-Y., and Kaufmann, K. (2018). Architecture of gene regulatory networks controlling flower development in Arabidopsis thaliana. Nat. Commun. *9*, 4534.
11. Varala, K., Marshall-Colón, A., Cirrone, J., Brooks, M.D., Pasquino, A.V., Léran, S., Mittal, S., Rock, T.M., Edwards, M.B., Kim, G.J., et al. (2018). Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants. Proc. Natl. Acad. Sci. USA *115*, 6494–6499.

12. Soyk, S., Lemmon, Z.H., Sedlazeck, F.J., Jiménez-Gómez, J.M., Alonge, M., Hutton, S.F., Van Eck, J., Schatz, M.C., and Lippman, Z.B. (2019). Duplication of a domestication locus neutralized a cryptic variant that caused a breeding barrier in tomato. Nat. Plants *5*, 471–479.

13. Hufford, M.B., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J.-M., Cartwright, R.A., Elshire, R.J., Glaubitz, J.C., Guill, K.E., Kaeppler, S.M., et al. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. *44*, 808–811.

14. Arnold, C.D., Nemčko, F., Woodfin, A.R., Wienerroither, S., Vlasova, A., Schleiffer, A., Pagani, M., Rath, M., and Stark, A. (2018). A high-throughput method to identify trans-activation domains within transcription factor sequences. EMBO J. *37*, e98896.

15. Tuttle, L.M., Pacheco, D., Warfield, L., Luo, J., Ranish, J., Hahn, S., and Klevit, R.E. (2018). Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex. Cell Rep. *22*, 3251–3264.

16. Belcher, M.S., Vuu, K.M., Zhou, A., Mansoori, N., Agosto Ramos, A., Thompson, M.G., Scheller, H.V., Loqué, D., and Shih, P.M. (2020). Design of orthogonal regulatory systems for modulating gene expression in plants. Nat. Chem. Biol. *16*, 857–865.

17. Soto, L.F., Li, Z., Santoso, C.S., Berenson, A., Ho, I., Shen, V.X., Yuan, S., and Fuxman Bass, J.I. (2022). Compendium of human transcription factor effector domains. Mol. Cell *82*, 514–526.

18. Alon, U. (2007). Network motifs: theory and experimental approaches. Nat. Rev. Genet. *8*, 450–461.

19. Thieffry, D., Huerta, A.M., Pérez-Rueda, E., and Collado-Vides, J. (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. Bioessays *20*, 433–440.

20. Rosenfeld, N., Elowitz, M.B., and Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. J. Mol. Biol. *323*, 785–793.

21. Song, L., Huang, S.-S.C., Wise, A., Castanon, R., Nery, J.R., Chen, H., Watanabe, M., Thomas, J., Bar-Joseph, Z., and Ecker, J.R. (2016). A transcription factor hierarchy defines an environmental stress response network. Science *354*, aag1550.

22. Konishi, M., and Yanagisawa, S. (2013). Arabidopsis NIN-like transcription factors have a central role in nitrate signalling. Nat. Commun. *4*, 1617.

23. Brooks, M.D., Cirrone, J., Pasquino, A.V., Alvarez, J.M., Swift, J., Mittal, S., Juang, C.L., Varala, K., Gutiérrez, R.A., Krouk, G., et al. (2019). Network Walking charts transcriptional dynamics of nitrogen signaling by integrating validated and predicted genome-wide interactions. Nat. Commun. *10*, 1569.

24. Krouk, G., Mirowski, P., LeCun, Y., Shasha, D.E., and Coruzzi, G.M. (2010). Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate. Genome Biol. *11*, R123.

25. Campbell, M.E., Palfreyman, J.W., and Preston, C.M. (1984). Identification of herpes simplex virus DNA sequences which encode a trans-acting polypeptide responsible for stimulation of immediate early transcription. J. Mol. Biol. *180*, 1–19.

26. Cress, W.D., and Triezenberg, S.J. (1991). Critical structural elements of the VP16 transcriptional activation domain. Science *251*, 87–90.

27. Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

28. Pan, C., Wu, X., Markel, K., Malzahn, A.A., Kundagrami, N., Sretenovic, S., Zhang, Y., Cheng, Y., Shih, P.M., and Qi, Y. (2021). CRISPR-Act3.0 for highly efficient multiplexed gene activation in plants. Nat. Plants *7*, 942–953.

29. Perez-Pinera, P., Ousterout, D.G., Brunger, J.M., Farin, A.M., Glass, K.A., Guilak, F., Crawford, G.E., Hartemink, A.J., and Gersbach, C.A. (2013). Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nat. Methods *10*, 239–242.

30. Yan, H., Pei, X., Zhang, H., Li, X., Zhang, X., Zhao, M., Chiang, V.L., Sederoff, R.R., and Zhao, X. (2021). MYB-Mediated Regulation of Anthocyanin Biosynthesis. Int. J. Mol. Sci. *22*, 3103.

31. Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell *12*, 2383–2394.

32. Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J., et al. (2016). Comparison of Cas9 activators in multiple species. Nat. Methods *13*, 563–567.

33. Lowder, L.G., Zhang, D., Baltes, N.J., Paul, J.W., Tang, X., Zheng, X., Voytas, D.F., Hsieh, T.F., Zhang, Y., and Qi, Y. (2015). A crispr/cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. Plant Physiol. *169*, 971–985.

34. Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646.

35. Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. *31*, 833–838.

36. Louvion, J.F., Havaux-Copf, B., and Picard, D. (1993). Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. Gene *131*, 129–134.

37. Ricci-Tam, C., Ben-Zion, I., Wang, J., Palme, J., Li, A., Savir, Y., and Springer, M. (2021). Decoupling transcription factor expression and activity enables dimmer switch gene regulation. Science *372*, 292–295.

38. Okamuro, J.K., Caster, B., Villarroel, R., Van Montagu, M., and Jofuku, K.D. (1997). The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. Proc. Natl. Acad. Sci. USA *94*, 7076–7081.

39. Erijman, A., Kozlowski, L., Sohrabi-Jahromi, S., Fishburn, J., Warfield, L., Schreiber, J., Noble, W.S., Söding, J., and Hahn, S. (2020). A High-Throughput Screen for Transcription Activation Domains Reveals Their Sequence Features and Permits Prediction by Deep Learning. Mol. Cell *78*, 890–902.e6.

40. Staller, M.V., Ramirez, E., Kotha, S.R., Holehouse, A.S., Pappu, R.V., and Cohen, B.A. (2022). Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains. Cell Syst. *13*, 334–345.e5.

41. Hope, I.A., Mahadevan, S., and Struhl, K. (1988). Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. Nature *333*, 635–640.

42. Jackson, B.M., Drysdale, C.M., Natarajan, K., and Hinnebusch, A.G. (1996). Identification of seven hydrophobic clusters in GCN4 making redundant contributions to transcriptional activation. Mol. Cell Biol. *16*, 5557–5571.

43. Tsugama, D., Yoon, H.S., Fujino, K., Liu, S., and Takano, T. (2019). Protein phosphatase 2A regulates the nuclear accumulation of the Arabidopsis bZIP protein VIP1 under hypo-osmotic stress. J. Exp. Bot. *70*, 6101–6112.

44. Liu, Y., Li, X., Li, K., Liu, H., and Lin, C. (2013). Multiple bHLH proteins form heterodimers to mediate CRY2-dependent regulation of flowering-time in Arabidopsis. PLoS Genet. *9*, e1003861.

45. Ikeda, M., Mitsuda, N., and Ohme-Takagi, M. (2009). Arabidopsis WUSCHEL is a bifunctional transcription factor that acts as a repressor in stem cell regulation and as an activator in floral patterning. Plant Cell *21*, 3493–3505.

46. Sigrist, C.J.A., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013). New and continuing developments at PROSITE. Nucleic Acids Res. *41*, D344–D347.

47. Tang, X., Lowder, L.G., Zhang, T., Malzahn, A.A., Zheng, X., Voytas, D.F., Zhong, Z., Chen, Y., Ren, Q., Li, Q., et al. (2017). A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. Nat. Plants *3*, 17018.

48. Ralser, M., Kuhl, H., Ralser, M., Werber, M., Lehrach, H., Breitenbach, M., and Timmermann, B. (2012). The Saccharomyces cerevisiae W303-K6001 cross-platform genome sequence: insights into ancestry and physiology of a laboratory mutt. Open Biol. *2*, 120093.

49. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. *47*, W191–W198.

50. 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.

51. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

52. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

53. Buchan, D.W.A., and Jones, D.T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res. *47*, W402–W407.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Agrobacterium tumefaciens strain GV3101	Joint BioEnergy Institute	N/A
Chemicals, peptides, and recombinant proteins		
Acetosyringone	Sigma-Aldrich	D134406
Critical commercial assays		
SsoAdvanced Universal SYBR Green Supermix	BioRad	#1725270
EZNA Plant RNA Kit	Omega Biotek	R6827
Experimental models: Organisms/strains		
Nicotiana benthamiana	JBEI	N/A
Saccharomyces cerevisiae: Strain background: W303	https://registry.jbei.org	JBx_099173
Oligonucleotides		
See Table S8		N/A
Recombinant DNA		
pYPQ152	Lowder et al.[33]	Addgene plasmid #69303
pYPQ131A	Lowder et al.[33]	Addgene plasmid #69273
pYPQ132A	Lowder et al.[33]	Addgene plasmid #69274
pYPQ133A	Lowder et al.[33]	Addgene plasmid #69275
pYPQ202	Tang et al.[47]	Addgene plasmid #86198
pms6370	https://registry.jbei.org	JBx_082980
Software and algorithms		
Cytoscape v3.9.0	Shannon et al.[50]	https://cytoscape.org
g:Profiler (Python)	Raudvere et al.[49]	https://biit.cs.ut.ee/gprofiler/page/apis
PSIPRED (Python)	Buchan et al.[53]	http://bioinf.cs.ucl.ac.uk/psipred/
Limma (R)	Ritchie et al.[51]	https://bioconductor.org/packages/release/bioc/html/limma.html
DESeq2 (R)	Love et al.[52]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Other		
Temporal RNA-seq data of *A. thaliana* nitrogen response	Varala et al.[11]	https://datadryad.org/stash/dataset/https://doi.org/10.5061/dryad.248g184
DAP-seq annotated genomic targets of TFs	O'Malley et al.[2]	http://neomorph.salk.edu/PlantCistromeDB

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Patrick M. Shih (pmshih@berkeley.edu).

Materials availability

All plasmid materials and bacterial strains will be made available through the Inventory of Composable Elements (https://public-registry.jbei.org/login). Sequences and raw data are available as supplementary materials.

Data and code availability

- All data are available in the supplemental information. All raw data from the parallel characterization of synthetic TFs are listed in Table S2. Protein sequences, normalized fold changes and general info of all TEDs are listed in Table S1. The annotated GRN is represented in Table S4.

- This paper does not report original code.
- This paper analyzes existing, publicly available data. The accession numbers for the datasets are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

N. benthamiana growth conditions

Wild type *N. benthamiana* plants were obtained from the in-house seed bank at the Joint BioEnergy institute. *N. benthamiana* plants were grown in SunGro Horticulture Professional Growing Mix #1 for four weeks in Percival-Scientific growth chambers at 25°C in 16/8-h light/dark cycles and 60% humidity at ~100 μmol of photons m^{-2} s^{-1}. Plants were fertilized two weeks after germination with MiracleGro. Post infiltration *N. benthamiana* plants were maintained in the same growth conditions.

Bacterial and yeast growth conditions

Agrobacterium tumefaciens strain GV3101 was obtained from the Inventory of Composable Elements (ICE) at the Joint BioEnergy Institute. Generated binary vectors were transformed into *A. tumefaciens* strain GV3101 and selected on LB plates (50 μg/mL kanamycin, 30 μg/mL gentamicin and 100 μg/mL rifampicin). Selected transformants were inoculated in liquid LB media with the same antibiotic concentrations. Yeast strain W303a (*MATa/MATα {leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15} [phi+]*) was obtained from the Inventory of Composable Elements (ICE) at the Joint BioEnergy Institute.

Strains were grown in synthetic complete glucose or galactose media without URA at 30°C. Overnight cultures were diluted (1:5) in the same respective media and grown for 4 h before flow cytometry.

METHOD DETAILS

Design of TED candidates

The candidate TF sequences were obtained from the work by O'Malley et al.[2] The DBDs of each candidate were identified using ScanProsite.[46] The aim was to extract the longest non DBD part of the TF that could have TED activity. Hence, in case of C- or N-terminal localization the DBD was removed from the TF sequence leaving a putative TF TED candidate. In case of DBD localization in the center of the protein the longest remaining TF TED candidate after truncation was chosen. An overview of the TED and DBD localization in each TF can be found in Fig. S1 and the exact location of the DBD and the motif is summarized in Table S1.

Construct design and assembly

The library of 529 TFs was obtained from O'Malley et al. and cloned into the binary vector pms7997 using Golden Gate cloning and construct specific primers (Table S8).[2] Binary vector pms7997 contained the plant codon optimized GAL4 DBD (amino acid 1–147) fused to the SV40 nuclear localization signal (PKKKRKV) and the GGSGG linker peptide connecting the DBD to individual TEDs. The synthetic TFs were driven by the MAS promoter with tNOS as a terminator. Plasmid assemblies were transformed into *Escherichia coli* strain DH5α and purified plasmids verified via sanger sequencing using primers pms7997_insertseq_fwd & pms7997_insertseq_rev. Using this approach, we were able to clone 403 putative TEDs. The PAP1-activator fusion constructs were assembled using golden gate cloning into vector pms057 with full-length PAP1 amplified from *A. thaliana* genomic DNA. Fusions of TEDs with dCas were generated by replacing VP64 in vector pYPQ152 using gibson assembly and otherwise assembled as described.[33,47] All vectors used for yeast experiments were generated using Gibson assembly of backbone pAI9 which contained the native yeast GAL4-DBD (amino acid 1–147) the SV40 nuclear localization signal (PKKKRKV) and the GGSGG linker peptide connecting the DBD to individual TEDs. The precise genomic locations of the native promoters NR1 and NIR1 cloned in front of GFP were chr1:29239368-29241368(−) and chr2:6808551-6810551(+) as obtained from TAIR, respectively. Full-length TFs were cloned into pms057 under the control of the 35S promoter. All primers used in this study are summarized in Table S8. All strains and plasmid maps are available in the Inventory of Composable Elements (ICE) at https://public-registry.jbei.org/login.

Agro-mediated transient transformation of *N. benthamiana*

Generated binary vectors were transformed into *A. tumefaciens* strain GV3101. Selected transformants were inoculated in liquid media with appropriate selection the night before the experiment. *A. tumefaciens* strains were grown until OD_{600} between 0.8 and 1.2 and were mixed equally (final OD_{600} = 0.5 for each strain) with the strain harboring the assay reporter construct to a final OD_{600} = 1.0. Cultures were centrifuged for 10 min at 4000 g and resuspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES, and 200 μM acetosyringone, pH 5.6). Cultures were induced for 2 h at room temperature on a rocking shaker. Leaves 6 and 7 of 4-week-old *N. benthamiana* plants were syringe infiltrated with the *A. tumefaciens* suspensions. Post infiltration *N. benthamiana* plants were maintained in the same growth conditions as described above. Leaves were harvested three days post infiltration and 16 leaf disks from two leaves per construct were collected. The leaf disks were floated on 200 μL of water in 96 well microtiter plates and GFP (Ex.λ = 488 nm, Em.λ = 520 nm) and RFP (Ex.λ = 532 nm, Em.λ = 580 nm) fluorescence measured using a Synergy 4 microplate reader (Bio-tek). The reporter construct for the screen was pms6370 containing GFP and dsRed expression cassettes. GFP expression was

driven by a fusion of five previously characterized GAL4 binding sites with the core WUSCHEL promoter.[16] GFP expression was normalized using dsRed driven by the MAS promoter on the same plasmid.

Quantification of anthocyanin content

Anthocyanin production experiments in *N. benthamiana* plants were performed as described above with the divergence that the entire infiltrated leaf tissue was collected from 2 infiltrated leaves per replicate. Collected tissue was flash frozen in liquid nitrogen and freeze dried at −50°C in vacuum for 24 h. The dried tissue was ground using bead beating for 5 min at 30 hz and 50 mg tissue was used for extraction. Anthocyanin was extracted three times using 1% hydrochloric acid in methanol and chlorophyll was removed with aqueous chloroform. Anthocyanin content was quantified by measuring absorbance at 535 nm on a Spectronic™ 200 spectrophotometer (Thermo Fisher Scientific).

Quantitative reverse transcription PCR experiments

Primers targeting the GUS and Kan genes were designed using the PrimerQuest software (IDT) (Table S8) and pre-screened for target specificity via Primer-BLAST against the *N. benthamiana* and *A. thaliana* genomes. qRT-PCR experiments were conducted on a BioRad CFX 96-well instrument using SYBR Green (BioRad). Reaction conditions were 1x ssoAdvance SYBR Green Supermix (BioRad) and 500 nM primers in 20 μL reactions, qPCR cycling parameters were 95 °C for 3 min, followed by 40 cycles of 30 s at 95 °C and 45 s at 56 °C. The linear dynamic range and efficiency of every primer set was verified over 1 x 10^2 to 10^9 copies per μL plasmid template. Target specificity was experimentally validated via melting temperature analysis.

For total RNA isolation, ~75 mg of leaf tissue was harvested from three plants 5 days post-infiltration, where one-half of the leaf was treated with reporter alone as reference and the other half with reporter and dCas9-TED candidate as the sample. Leaf tissue was flash frozen in liquid nitrogen and RNA extracted using the EZNA Plant RNA Kit I (Omega Biotek). DNA contamination was removed by treating total RNA with Turbo DNase with inactivation reagent (Invitrogen). cDNA was generated from 1.0 μg total RNA using SuperScript IV Vilo reverse transcriptase (Thermo Fisher Scientific). qRT-PCR was carried out using 1 μL of the reverse transcription reaction as a template. For all experiments, a no template- and a no reverse transcription control was run. All primers were tested with wild type cDNA from plant tissue treated with *Agrobacterium* (strain GV3101) containing an empty vector control with Cq > 36 as the threshold for no off-target activity. The ΔΔCq method was used to determine normalized expression with GUS as the sample- and KanR as the reference gene quantified.

Flow cytometry

For experiments in *S. cerevisiae* lab strain W303a (*MATa/MATα {leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15} [phi+]*) was used.[48] The GAL1-GFP reporter cassette was integrated into the URA3 locus. The native Gal4-TED fusions were expressed using the TEF1 promoter in a 2μm-plasmid in the reporter strain. For flow cytometry experiments all strains were grown in CSM-URA (Sunrise Science Products) media prepared following the suppliers manual with 2% w/v Glucose, except for the positive control which was grown in 2% w/v Galactose. Experiments were performed on the BD Accuri™ C6 flow cytometer (BD Biosciences), samples were washed with cold 1x PBS (137 mmol NaCl, 2.7 mM KCl, 1.8 mM KH_2PO_4, 10 mM Na_2HPO_4) once before measurement in 1x PBS. Per sample 100,000 events were recorded and analyzed using the FlowJo™ software.

QUANTIFICATION AND STATISTICAL ANALYSIS

Grouping of TED populations

We established the general repressor and activator populations by first grouping all TEDs into two groups based on decreased or increased gene expression compared to the reporter only control. We then individually defined the repressor or activator population using a Kruskal-Walis test paired with a Dunnets ad hoc test. The first 10 consecutive TEDs significantly differing from the reporter measurements ($p < 0.05$) marked the boundaries of the populations which set the threshold of the repressor and activator population around −0.68 and 1.00 $\log_2$ Normalized Fold change, respectively (Table S1). We further sub-grouped the activator population by defining strong activators as TEDs with >2 $\log_2$ Normalized Fold change.

Gene ontology enrichment of activators and autoregulation

DNA binding targets of TFs in this study were obtained from the Arabidopsis DAP-seq database (http://neomorph.salk.edu/PlantCistromeDB).[2] GO term enrichment of the target genes of TFs screened in this study was performed using the g:Profiler web service accessed via the Python API[49] with the datasource limited to GO:biological process and the significance threshold method set to default (g_SCS). To study strong activators we focused on TFs whose TEDs increased gene expression more than VP16. As many TFs target the promoters of other TFs we excluded GO terms linked to transcription. GO terms inside the same functional class were manually sorted.

To study autoregulation we assigned a Boolean value to every TF, whose TED we studied here, based on whether it had the potential to bind its own promoter region. The Boolean value 1 was assigned to TFs with binding and 0 to TFs with no binding to their own promoter region. We then grouped the Boolean values into the TED populations and studied a potential auto-regulation bias between TED populations using a two-sided Fisher's exact test.

Generating a *trans*-regulatory annotated nitrogen response GRN

The extended nitrogen response GRN was built on a version including DNA binding information and a co-expression machine learning model based on temporal RNA-seq data.[11] In short, Varala et al. performed a temporal RNA-seq study of the response of *A. thaliana* to nitrogen feeding after nitrogen starvation. They collected tissue of plants at different time points post nitrogen feeding as well as an untreated control. By comparing the RNA-seq profile of the induced and untreated samples they established a set of nitrogen responsive genes, defined as the first time point where expression in the induced sample was $\geq$1.5-fold of control. They further generated a pruned GRN only including TF-Gene interaction that had previously been verified *in vitro*. We utilized this network to annotate TF-Gene interactions with TED data, represented as an edge attribute. We visualized the GRN and extracted TF-gene interactions of NR1 and NIR1 using Cytoscape v3.9.0.[50] To study the effects of the activator group observed at 15 min we performed RNA-seq analysis using the limma package and DESeq2 in R as shown in Varala et al.[51,52] Rate of induction of gene expression was calculated by subtracting Fold changes of consecutive time points. The time point of maximal induction and maximal decrease of gene expression was derived from these subtractions.

Validation of GRN predicted TF-Gene interactions

To verify regulatory activity of TFs binding NR1 and NIR1 promoter regions in the nitrogen GRN we utilized Agro-mediated transient transformation as described above. In this setup we used three *Agrobacterium* strains expressing one native full-length TF, a GFP reporter driven by the respective promoter regions of NR1 and NIR1 and mScarlet driven by the NOS promoter for normalization. Modulation of normalized GFP expression from basal expression was validated using the Mann-Whitney-U test on n = 40 replicates and statistical significance defined as $p < 0.05$.

Localization of activation domains in plant TFs using ADpred

We localized putative activation domains in the TEDs from our study using the ADpred model.[39] The model analyzes sequence stretches of 30 amino acids with predicted secondary structure information. Therefore, the secondary structure of full-length TED domains was predicted using the PsiPred workbench.[53] For the analysis of individual TEDs, we fragmented the protein sequence into 30 amino acid sections moving 5 amino acids in between fragments. If any of the fragments of a given TED scored at $\geq$ 0.9 in the ADpred model the TED potentially contained an AD. We assigned a Boolean to every TED based on the scoring, 0 for no AD and 1 for containing a potential AD. A potential bias between different TED populations was observed using a two-sided Fisher's exact test. To validate the predictive power of the model we chose 3 strong activators and subcloned the longest consecutive amino acid stretches scoring above the ADpred threshold of 0.9 into pms7997 and assayed their performance in comparison to their TED counterparts in *N. benthamiana*. Modulation of normalized GFP expression from basal expression was validated using the Mann-Whitney-U test on n = 16 replicates and statistical significance defined as $p < 0.05$.