Optimized dCas9 Programmable Transcription Activators for Plants

Matthew H Zinselmeier[1,3,4], **J. Armando Casas-Mollano**[2,3,4], **Adam Sychla**[2,3,4], **Stephen C Heinsch**[3,4,5], **Daniel F Voytas**[1,3,4,✉], **and Michael J Smanski**[2,3,4,5✉]

[1]Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, MN 55455
[2]Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, MN, 55455
[3]Center for Precision Plant Genomics, 1500 Gortner Avenue, Cargill Building, Saint Paul, MN 55108
[4]Biotechnology Institute, University of Minnesota, Saint Paul, MN 55108
[5]Bioinformatics and Computational Biology Graduate Program, University of Minnesota, Saint Paul, MN 55108

Understanding how the expression of genes impacts plant development and physiology is important for rationally engineering crop improvements. Programmable Transcription Activators (PTAs), including CRISPRa activators, have traditionally relied on a limited number of transcription activation domains, namely the VP64 domain derived from human herpes simplex virus, to control gene expression. We reasoned there was considerable space for PTA improvement by replacing this domain with a plant-derived activation domain. To address this, we designed, built, and tested a PTA library of 38 putative plant transcription activation domains. Domains from HSFA6b, AvrXa10, DOF1, DREB1, and DREB2 genes function as strong activators in *Setaria viridis* and *Arabidopsis thaliana* both in protoplast assays and in transgenic plants. Overexpression of multiple endogenous genes (*FT, PAP1, WUS*) reached levels similar to the highly expressed housekeeping gene, PP2A, regardless of basal expression level. Further, these domains were effective in different PTA architectures, including the dCas9-SunTag, dCas9-Moontag, and TALE-SunTag systems. Lastly, we demonstrate the ability of these improved PTAs to map enhancer regions that promote gene expression in plants. This work showcases the effective and flexible nature of PTAs to activate target genes in plants, providing tools that can be used to improve agronomically relevant traits of interest.

dCas9 | Transcription activator | Plant engineering | SunTag
Correspondence: ***voytas@umn.edu, smanski@umn.edu***

DRAFT

Introduction

Controlling the expression of endogenous genes in plants is important for basic research into plant development. Plant genomes provide the blueprint for plant growth, survival, and reproduction. In order to survive a plant must also respond to surrounding environmental conditions such as temperature, light, or humidity. Gene expression is the process by which the genetic blueprint is put into action; activating genes at different times and places to develop, survive, and reproduce in the face of external stimuli (1, 2). Studies correlating differential gene expression with phenotype allows for predicting which genes are most important for proper development and responding to the environment (3, 4). The ability to control the spatio-temporal expression of these key genes driving phenotypic variation will be critical for trait and yield improvement of crop species in different environments across the world.

To this end, recent advances have led to the development of new tools capable of increasing gene expression. Programmable Transcription Activators (PTAs) are fusion proteins comprised of a transcription activation domain (AD) fused to a DNA-binding domain that can be rationally engineered to recognize a DNA sequence of interest. The DNA-binding domain can be a zinc-finger (cite), a transcription activator-like effector (TALE) DNA-binding domain (cite), or a catalytically 'dead' Cas9 protein (dCas9) that is directed to a target sequence via a single guide RNA (sgRNA) (5). PTAs can drive the over- or ectopic-expression of endogenous genes when designed to bind to a promoter region (6, 7). The strength of overexpression is strongly correlated to a target gene's basal expression levels; PTAs targeting lowly-expressed genes can achieve higher fold-overexpression values than targeting highly-expressed genes (8). dCas9-based PTAs can target many promoters in parallel by co-expressing multiple sgRNAs (9). PTAs have been used to explore gene function and engineer barriers to sexual reproduction in insects (10).

The first dCas9-based PTAs utilized a VP64 AD. VP64 is a tetrameric repeat of the VP16 protein derived from herpes simplex virus (11, 12). This dCas9-VP64 design has been used plants to activate target genes (13). An additional improvement to PTAs came with dCas9 translational fusion to an AD termed VPR (14). VPR is a translational fusion of three different ADs - VP64, RTA, a p65. While VPR drives strong gene expression in animal species, the same robust effectiveness seen in mammalian cells remains elusive in plants (5).

Several groups have developed improved dCas9-based PTAs for use in plant systems. The most common strategy is to increase the number of ADs recruited to the PTA. dCas9-TV is comprised of 6 copies of the TAL activation domain and two copies of the VP64 activation domain translationally fused to dCas9 (15). The SunTag system uses a non-covalent interaction between single-chain variable fragment antibodies (scFv) and a tandemly-repeated epitope tail (GCN4 motif) to recruit ADs to the dCas9 (16). Using 10 copies of GCN4 fused to dCas9, the SunTag system can activate target genes in *Arabidopsis thaliana* to produce visible phenotypes (17). While these PTAs improved performance in plants, they still rely on the VP64 AD. Given the evolutionary distance between mammals and plants, we sought to improve PTA activity by screening ADs that evolved in plant systems. Here we

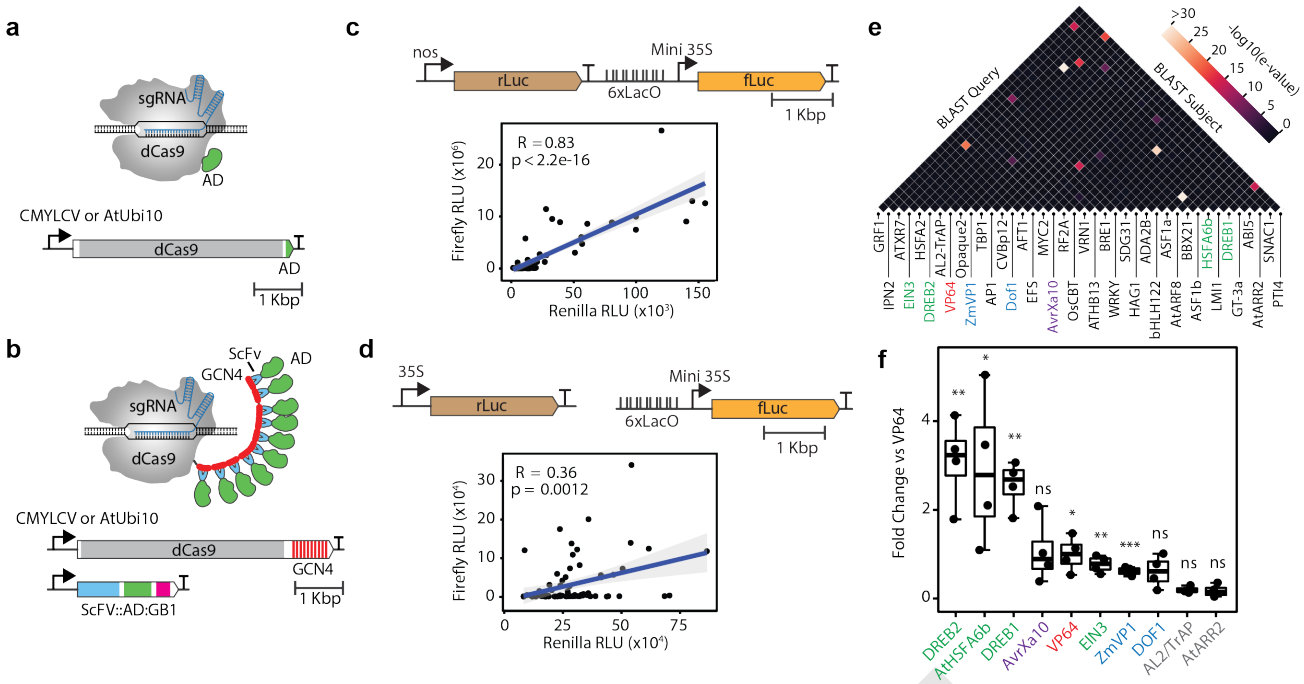


Fig. 1. Screening a library of plant-derived activation domains with the dual luciferase protoplast assay. Cartoon images of dCas9-based Programmable Transcription Activators (PTAs) including the direct fusion (a) and SunTag scaffolding system (b). (c) Genetic construct design for a single-plasmid dual luciferase assay (top) and observed correlation between expression of PTA-controlled Firefly luciferase and constitutively expressed Renilla luciferase (bottom). Data is from *S. viridis* protoplasts. (d) Genetic construct design for two-plasmid dual luciferase assay (top) and corresponding low correlation between Firefly and Renilla luciferase expression (bottom). Data is from *A. thaliana* protoplasts. (d) Heatmap showing the similarity of tested activation domains. Black denotes no significant similarity. Color-coded ADs highlight the strongest from the analysis in (f), with the following relationship: green, dicot; blue, monocot; purple, bacterial; red, human virus. (f) A rank order of the strongest ADs is show for *A. thaliana* protoplasts against the VP64 positive control. Parametric t-tests were performed comparing each AD with the negative control (-)sgRNA. * corresponds to p <= 0.05, ** corresponds to p <= 0.01, *** corresponds to p <=0.001, and **** corresponds to p<= 0.0001.

report the discovery and performance of several plant-derived ADs for programmable gene expression in monocot and dicot plants.

Results

Building a library of putative plant-derived transcription activation domains. We first determined a suitable experimental system for identifying strong ADs. Protoplast isolation and transformation pipelines yield consistent cell numbers and transformation efficiencies, allowing for direct comparison of groups within a given transformation when activating a reporter gene (18, 19).

We compared the ability of direct-fusion PTA architectures such dCas9-VP64 with TAD-scaffolding PTA architectures like SunTag (Figure 1a-b) to activate an endogenous target gene in *A. thaliana*. A tRNA-based multi-guide expression array (20) expressing four sgRNAs targeting a single core promoter for *WUSCHEL (WUS)* was transformed into *A. thaliana* protoplasts along with a dCas9 PTA, followed by RNA extraction and RT-qPCR. The dCas9-PTAs were expressed using the *Ubi10* promoter from *A. thaliana*.

At one day post-transformation, all PTAs increased the expression of *WUS* compared to a no-sgRNA control. While dCas9-TV outperformed dCas9-VP64, the SunTag-VP64 design outperformed both direct-fusion PTAs by two orders of magnitude (Supplemental Figure 1). Given these results, we reasoned that testing a library of putative ADs as fusions to scFv in the SunTag system would result in a dynamic range of activation for comparison.

We compiled a list of ADs to test in the context of plant PTA applications. We started with a literature search of transcription factors with known DNA binding domains from diverse protein families. We computationally removed native DNA-binding domains, as these are expected to produce off-target effects if retained in PTAs. If an activation domain was not empirically determined, we selected motifs enriched in acidic and/or aromatic residues as these patches are often associated with transcription activation domains due to their propensity to form phase separation condensates upon recruitment to a core promoter (21). We also added activation domain sequences from plant pathogen effector proteins, such as TALE proteins from *Xanthomonas*, along with sequences derived from transcription preinitiation complexes such as 14-3-3 scaffolding proteins.

The name, sequence, and citations for each domain tested can be found in Supplementary Table S1. We observe low amino acid sequence similarity across coding sequences from which ADs were derived, as illustrated by an all-by-all BLAST analysis (Figure 1e). This highlights the aim of this screen to survey a diverse array of plant-related ADs, versus optimizing on a well-performing family.

Putative AD sequences were codon optimized based on average codon usage tables for *Arabidopsis thaliana* and *Setaria viridis* for reliable expression across a variety of plant backgrounds and synthesized as dsDNA fragments. We used Type IIS restriction enzyme cloning to assembly putative AD

DRAFT

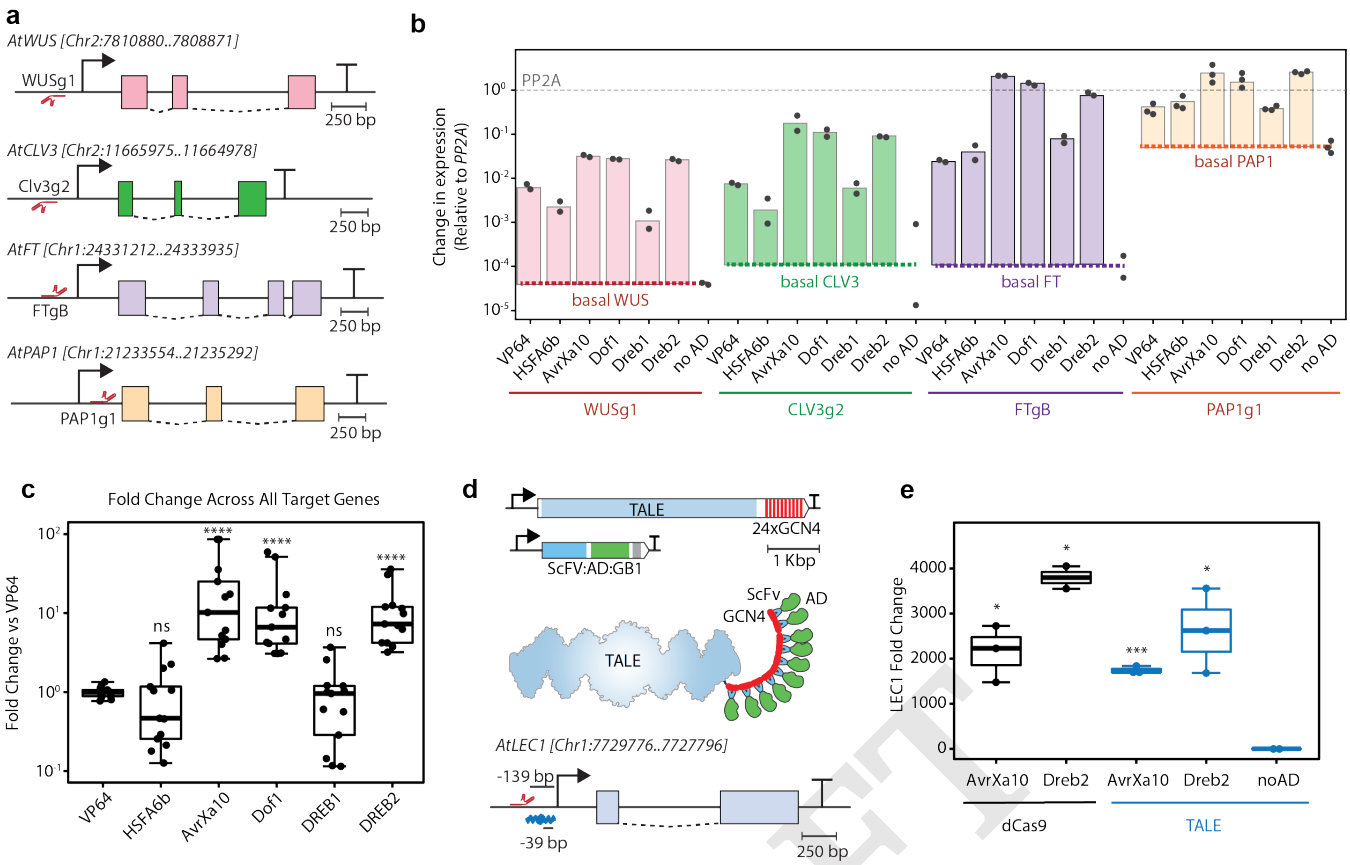


Fig. 2. Endogenous gene activation in *A. thaliana* protoplasts. (a) Guide RNA design for a given promoter is shown with sgRNA distance to the known TSS for *WUS*, *CLV3*, *FT*, and *PAP1* loci. (b) Gene activation data for each corresponding gene is shown relative to the housekeeping gene *PP2A*. Basal level of gene expression, corresponding to values measured from no AD controls is unique for each target and is highlighted with colored, dashed line. (c) Aggregate gene activation across all genes tested, where each dot represents an independent protoplast transformation. (d) The TALE DNA binding domain was fused to the GCN4 epitope tail to comprise a TALE-SunTag system (top). The TALE DNA binding domain was engineered to target the LEC1 promoter, compared to dCas9 binding the same promoter (bottom).(e) The *LEC1* target gene in *A. thaliana* was targeted for gene activation comparing dCas9-SunTag (black) with TALE-SunTag (blue). Parametric t-tests were performed comparing each AD with the negative control NoAD. * corresponds to p <= 0.05, ** corresponds to p <= 0.01, *** corresponds to p <=0.001, and **** corresponds to p<= 0.0001

coding sequences into an scFv destination vector. ADs were expressed as C-terminal fusions to scFv, separated by a flexible GS linker. ADs coding sequences that failed multiple attempts to assemble in the scFv vector were removed from the study.

A set of 7 plant-evolved ADs show comparable activity to VP64 in a dual luciferase protoplast assay. A reporter assay was designed to rapidly screen the library for domains capable of activating transcription in plant cells based on the dual luciferase assay (22, 23). On the reporter plasmid, we designed a synthetic promoter comprising six copies of the *Lac* operator (*Lac*O) upstream of the minimal 35S CaMV core promoter upstream of Firefly luciferase (24). An sgRNA targeting the *LacO* sequence results in dCas9 binding all six *LacO* motifs, such that firefly luciferase expression reports activation domain strength. To provide an internal control of plasmid copy number, a Renilla luciferase controlled by a constitutive promoter was inserted in the same plasmid, upstream of and oriented in the same direction as the Firefly luciferase (Figure 1c).

Alternative activation domains were expressed as scFv-AD fusions driven by a 35S promoter, allowing for strong expression in both monocot and dicot protoplasts. The dCas9-24xGCN plasmid was expressed by a *AtUBI*10 promoter in *A. thaliana* or the CmYLCV promoter in *S. viridis* (Figure 1a).

The library was first tested in *Setaria viridis* protoplasts. We isolated and transformed 100,000 cells with 2 micrograms of DNA in 96 well plate format. Each transformation received dCas9-24xGCN, the *LacO* sgRNA, the single dual luciferase plasmid with 35S driving Renilla, and a given scFv-AD. Each scFv-AD was independently transformed twice per protoplast isolation. Following two protoplast isolations spanning four independent transformations, we noticed that DREB2, DREB1, and HSFA6b consistently produced the largest Renilla luciferase values across all four independent transformations. While the first data set showed a weak correlation (R=0.15, Supplemental Figure 2), when fitting the second data set to a linear model we noticed a strong correlation between Firefly and Renilla luciferase activity (R=0.53, Supplemental Figure 3). This seemed unlikely given our initial construct design in which Renilla luciferase values should be a readout of transformation efficiency, and should not fit a model where firefly luciferase expression is linearly correlated with Renilla luciferase expression. We reasoned that

DRAFT

transcription activation machinery being recruited by the AD to the minimal 35S promoter may also activate transcription at the nearby upstream Renilla luciferase promoter. In particular, strong ADs may be activating nearby genes to a greater extent than weaker activation domains. If this was the case, driving the Renilla luciferase with a weaker promoter may result in a better linear fit due to strong ADs at the the firefly luciferase minimal promoter activating the nearby Renilla luciferase. To test this hypothesis of local off target activation, we repeated the library transformation with a weaker constitutive *Nos* promoter driving the Renilla luciferase. Indeed we observed a strong linear correlation between the Firefly and Renilla luciferase values when the upstream nearby Renilla luciferase was driven by the weak *Nos* promoter (Figure 1c).

Given these results, we decided to split the Firefly and Renilla luciferase to separate plasmids (Figure 1d). If the Firefly and Renilla luciferase promoters are on separate plasmids, we hypothesize that the likelihood of ADs at one promoter impacting the expression of a promoter on a separate plasmid should be low. We tested the split plasmid luciferase platform in *A. thaliana* protoplasts with a smaller library of ADs and saw a weak correlation for a linear regression model (Figure 1d, bottom). We performed RLU calculations using Renilla as a copy number control and observed a similar set of core set of ADs was capable of activating transcription in *A. thaliana* for this experiment, albeit with variable strength compared to the *S. viridis* data sets (Supplemental Figure 4). DREB2 retained high activity, and AvrXa10 performed better in *A. thaliana* than in *S. viridis*. Conversely, DREB1, HSFA6b, and ZmVP1 all displayed lower levels of activation in the dicot model *A. thaliana*.

With these results suggesting local off target activation at the Renilla luciferase promoter, we decided to display the *S. viridis* data as raw Firefly values compared to the VP64 positive control. When performing this analysis a core set of 7 activation domains including DREB2, HSFA6b, DREB1, AvrXa10, EIN3, ZmVP1, and DOF1 showed comparable or greater activity than the conventionally used VP64 activation domain (Figure 1f). Notably, DREB2, HSFA6b, and DREB1 showed the largest Firefly activity across four protoplast transformations. The full library data sets for both *A. thaliana* and *S. viridis* can be found in Supplemental Figures 5 and 6.

Finally, we wanted to see how our plant-derived ADs performed in a crop species. We tested another small library of strong plant ADs in the model crop *Zea mays* using the split dual luciferase assay. *Z. mays* protoplasts were isolated from greened tissue, and cells were transformed with AtHSFA6b, AvrXa10, DOF1, DREB1, DREB2, and VP64 ADs in the SunTag system. Both the dCas9-24xGCN and scFv-AD constructs were driven by the *ZmUbi1* promoter for strong expression. We performed duplicate luciferase expression analysis using Renilla luciferase as copy number control and saw both DREB2 and DOF1 retain strong activity in the model crop *Z. mays* (Supplemental Figure 7). While the average observed fold-change in gene expression was 3x greater for DREB2 compared to VP64, the experiment did not have the statistical power to determine if this effect size is significant. Interestingly, the AvrXa10 domain decreased in activity within *Z. mays* protoplast relative to its strength in *A. thaliana* protoplasts. Taken together, we conclude that HSFA6b, AvrXa10, DOF1, DREB1, and DREB2 show promising activity across all three species tested when compared to the standard VP64 AD.

DREB2, DOF1, and AvrXa10 plant-derived activation domains outperform VP64 across endogenous loci. We next tested the efficiency of HSFA6b, AvrXa10, DOF1, DREB1, DREB2, and VP64 at endogenous loci. Single guide RNAs expressed from the *A. thaliana* U6 promoter were constructed to target the core promoters of *FLOWERING LOCUS T* (*FT*), *CLAVATA3* (*Clv3*), *PRODUCTION OF ANTHOCYANIN PIGMENT 1* (*PAP1*), and *WUSCHEL* (*WUS*) for activation (Figure 2a-b, Supplemental Figures 6-8). All sgRNAs used in this study are previously published (13, 25), except for the *WUS* guide, which was developed for this study. *A. thaliana* protoplasts were isolated and transformed with the SunTag activator and U6-sgRNA, with either two or three replicates per activation domain per target gene. After 24 hours, RNA was isolated and RT-qPCR was performed to quantify gene activation. We compared the activity of each AD against a no AD control for every sgRNA tested, in which the dCas9 and sgRNA were transformed without an scFv-AD. For each sample, the target gene and a control housekeeping gene (*PP2A*) were amplified. We report expression relative to *PP2A* in Figure 2b for *WUS*, *CLV3*, *FT*, and *PAP1*. By graphing all data points relative to the housekeeping gene, it is easier to observe both the fold-change in gene expression (height of the bar), and the absolute change in gene expression (location relative to the y-axis). This is important, as the fold-change in expression that can be achieved by PTAs is inversely correlated with the basal expression level. Genes already expressed at high levels cannot be overexpressed to the same relative degree as genes with a low basal expression level(8).

The sgRNAs bind the core promoter regions, located 0-300 base pairs upstream of an annotated transcription start site (TSS) for the gene. The PAP1 sgRNA used in this study binds downstream of the annotated TSS. A clear pattern emerged across sgRNAs tested in which AvrXa10, DOF1, and DREB2 consistently produced greater gene activation than VP64, HSFA6b, and DREB1 (Figure 2b, Supplemental Figures 8-9). The rank order of AvrXa10, DOF1, and DREB2 changed depending on the guide being tested. As expected, the maximum fold-overexpression was correlated with basal expression level. For example, FT saw a much greater fold-overexpression compared to no-AD control (>10,000) than did PAP1 (50). However this is due to FT starting at much lower basal expression levels, and maximum absolute expression attained is comparable. WUS, on the other hand, could only be overexpressed to 10% the level of PP2A.

We compared each AD's gene activation across the panel of endogenous genes to VP64. After setting VP64 activation values for each gene tested equal to 1, we combined data sets for all sgRNAs across four target genes (n=13). The same

DRAFT

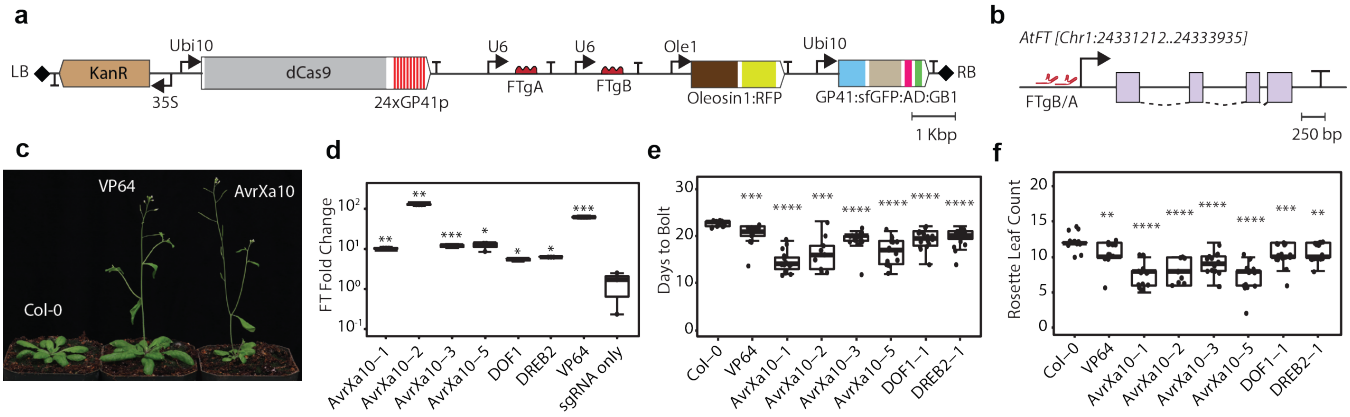


Fig. 3. FT gene activation in transgenic plants (a) Genetic construct design for binary vectors floral dipped into transgenic *A. thaliana* plants. (b) Illustration of sgRNA design relative to the FT TSS. (c) Representative images of T2 transgenic plants for Col-0, VP64, and AvrXa10 transgenic lines. (d) Molecular quantification (RT-qPCR) of FT gene activation from 10d old T2 seedlings. Statistical analysis was performed on triplicate technical replicates from six pooled plants from each genotype. (e) Bolting time quantification in T2 transgenic plants. (f) Rosette leaf quantification one day after bolting in T2 transgenic plants. Parametric t-tests were performed comparing each line with the negative control Col-0. * corresponds to p <= 0.05, ** corresponds to p <= 0.01, *** corresponds to p <=0.001, and **** corresponds to p<= 0.0001

trend was present in which AvrXa10, DOF1, and DREB2 produced statistically significant gene activation across all target genes and guide RNAs that were tested (Figure 2c).

We tested the portability of these ADs between different programmable DNA binding domains. We cloned the 24xGCN4 epitope tail from the SunTag system as a C-terminal fusion to a TALE programmable DNA binding domain. A TALE DNA binding domain is engineered by assembling a series of 20 TALE repeats, each of which contain a unique Repeat Variable Diresidue (RVDs)(13). Each RVD within a TALE repeat specifies a particular nucleotide for binding, such that a 20 base pair genomic target sequence is specified by a combination of 20 RVDs. We built TALE repeats targeting a 20 bp region 39 bp upstream of the transcription start site for the *A. thaliana* gene *LEC1*. In addition, we cloned an sgRNA targeting a 20 bp region 139 bp upstream of the transcription start site for the same *LEC1* target gene (Figure 2d). We then tested the ability of the AvrXa10 and DREB2 ADs to activate the *LEC1* target gene in conjunction with either dCas9 or TALE DNA binding domains in *A. thaliana*. As expected the dCas9 SunTag system using AvrXa10 and DREB2 domains resulted in strong *LEC1* gene expression with 2100 or 3800 fold greater activation than the negative control, respectively (Figure 2e). The TALE SunTag system using AvrXa10 and DREB2 domains also resulted in strong *LEC1* gene expression of 1700 and 2600 fold greater activation than the negative control, respectively. These results demonstrate the flexibility of the AvrXa10 and DREB2 ADs to drive strong expression of target genes using different programmable DNA binding domains.

Avrxa10, DOF1, and DREB2 activation domains promote early flowing in transgenic plants. Next, we tested whether our plant-derived ADs would produce a phenotype upon gene activation in stable transgenic plants. We generated binary vectors expressing our plant-derived ADs using a new PTA architecture named the MoonTag system (Casas-Mollano, in preparation). The MoonTag system utilizes an epitope-nanobody interaction to recruit the AD to the dCas9 molecule (26). There are 24 copies of the GP41p epitope fused to dCas9, while AvrXa10, DOF1, and DREB2 were fused to the GP41 2H10 nanobody. Two sgRNAs targeting the FT core promoter in *A. thaliana* were expressed from U6 promoters to drive FT gene activation (Figure 3a-b). Plants were floral dipped and T1 transgenic lines were identified using the pFAST Oleosin1 seed coat fluorescent reporter (27, 28). For comparison, we used a validated MoonTag-VP64 transgenic line as the positive control(29) and wild type Columbia-0 (Col-0) plants as the negative control in these experiments. In the T1 generation, many plants showed an early flowering phenotype (Supplemental Figure 9). We allowed these plants to set seed and used the homozygous plants in the T2 generation for quantification of flowering time.

Sets of 20 seedlings per line were planted on selection media and allowed to germinate. Ten days after germination, seedlings were pooled into groups of six and RNA was extracted from the pooled group. We measured FT gene expression in the transgenic lines compared to a control line expressing only sgRNAs against a different target gene. We observed strong gene activation in the positive control VP64 line, along with one of the AvrXa10 transgenic lines (Figure 3d).

We quantified the FT overexpression phenotype in the T2 population by selecting 18 RFP-positive seedlings and planting them directly in soil. Following germination, we waited for the seedlings to produce their first set of true leaves and used this time point as t=0. Two commonly used metrics of flowering time are bolting time and rosette leaf number (30, 31). We noted the day at which bolting was observed and calculated the difference between this day and t=0 to quantify bolting time (Figure 3e). We then counted rosette leaves one day post bolting for each seedling as an additional quantification of the FT phenotype (Figure 3f). We observed a strong FT activation phenotype across multiple AvrXa10 transgenic lines, with early bolting and reduced rosette leaf number in comparison to wild type plants. We also noted reduced rosette leaf size in several of the AvrXa10 plants (Figure 3c) compared to Col-0 and VP64 plants. We also

DRAFT

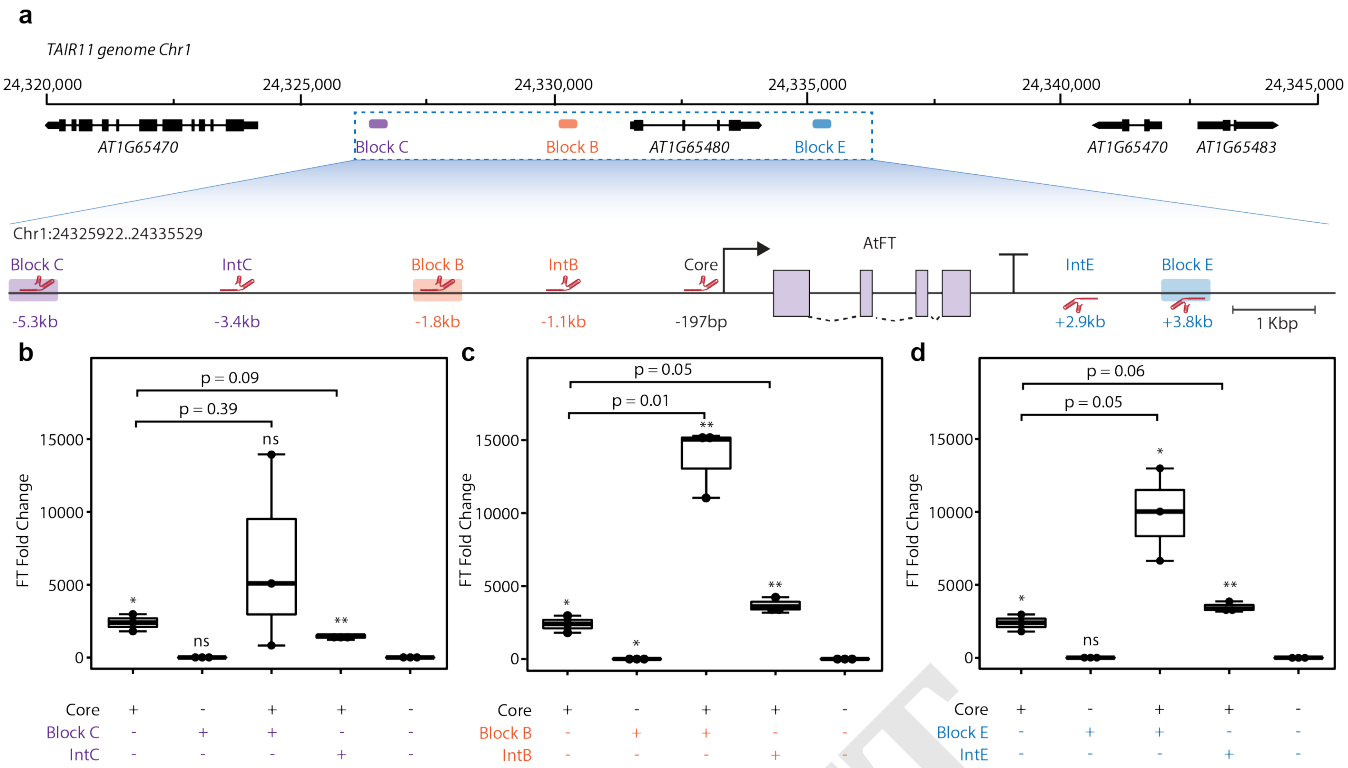


Fig. 4. PTA-mediated enhancer activation at the *Arabidopsis thaliana* FT locus. (a) Schematic illustration of sgRNA design in relation to the region of *A. thaliana* chromosome 1 surrounding the FT gene. Top depicts entire 25kb locus and bottom shows more detail of enhancer location and sgRNA design. sgRNAs drawn above or below the line denote targeting the sense or antisense strand, respectively. Distances labeled are relative to the FT TSS. (b-d) FT gene activation, measured by RT-qPCR when targeting Block C (b), Block B (c) or Block E (d) with different combinations of sgRNAs (x-axis). Parametric t-tests were performed comparing each AD with the negative control where no sgRNA is delivered. * corresponds to p <= 0.05, ** corresponds to p <= 0.01, *** corresponds to p <=0.001, and **** corresponds to p<= 0.0001

observed early flowering phenotypes for VP64, DOF1, and DREB2, albeit weaker than the AvrXa10 lines. Taken together we observed our plant-derived AvrXa10, DOF1, and DREB2 function in transgenic plants, with AvrXa10 showing the strongest gene activation and phenotype within the dicot species *Arabidopsis thaliana*.

PTA-Mediated Gene Enhancer Activation. Finally, we demonstrated the ability of our improved plant PTAs to map enhancers. Co-targeting PTAs to both a core promoter and enhancer region has previously been shown to boost observed expression levels in human cells (32). To illustrate this concept in plants, we utilized the *FT* gene activation platform. Of the two sgRNAs tested, FTgB was capable of driving stronger gene activation than FTgA at the core promoter in our protoplast assays (Figure 2b, Supplemental Figure 8). According to previous literature, Block B (-1.8kb), Block C (-5.3kb), and Block E (+3.8kb) are three known enhancers of FT located either upstream or downstream of the TSS (33). We designed sgRNAs that target the centers of each 350-400 bp enhancer, along with the FT core promoter guide FTgB previously tested. In addition, we designed guides that target sequences located between the annotated enhancer regions and the core promoter (Figure 4a).

For each of Blocks B, C, and E, targeting the enhancer alone did not result in overexpression above no-sgRNA controls (Figure 4b-d). Targeting both the core promoter and the enhancer led to a significant increase in expression level for Blocks B and E. The Block C enhancer-targeting sgRNA produced greater expression levels in two of three replicates, but a large amount of variance in this experiment prevented it from reaching statistical significance. When targeting enhancers plus core promoters, mean fold overexpression compared to no-sgRNA controls ranged from 6,000-fold (Block C) to 14,000 fold (Block B). This improvement over targeting the core promoter alone (2,400-fold overexpression compared to control) could not be attributed to having a second sgRNA, as the sgRNAs targeting sequences outside of the enhancer regions did not significantly boost expression levels (Figure 4b-d).

Discussion

We set out to identify plant-derived ADs that function as well or better than VP64 in plant systems. Through a library approach we were able to clone and test putative ADs in protoplast dual luciferase assays to determine which sequences showed activity comparable to VP64. We identified a core set of 5 activation domains comprised of AvrXa10, HSFA6b, DREB1, DREB2, and DOF1 that showed strong activity in both the monocot model *S. viridis* and the dicot model *A. thaliana* dual luciferase protoplast assays. The high throughput nature of protoplast transformation serves as a valuable tool for testing large numbers of genetic parts (19).

While performing these assays, we observed what appeared to be local off-target activation of our dual luciferase re-

DRAFT

porter. DREB2, DREB1, and HSFA6b consistently produced the largest Renilla luciferase values, even though they were targeted to the Firefly luciferase promoter. This resulted in a linear correlation between Firefly and Renilla values (Figure 1b). Because this correlation was abrogated by splitting the reporters into two plasmids (Figure 1c), we conclude expression from the upstream Renilla promoter is the result of trans-activation from the PTA. In this case, the promoter driving the Renilla gene was located approximately 3kb upstream of the sgRNA binding site. Interestingly, targeting a PTA to within 3kb of the endogenous FT locus did not cause FT overexpression unless at least one sgRNA bound to the core promoter region. It is possible that the proximity effect seen in our single-plasmid dual luciferase assay is dependent on transcriptional activity from the targeted promoter. In addition, the *A. thaliana* genome is organized into nucleosomes whereas the luciferase reporter plasmids do not contain nucleosomes. This could facilitate either contact between both promoters. This phenomena may vary depending on the locus, the number of sgRNAs, or even the AD that is directed to a target site. In genome-wide screens, off-target activation events are rarely seen (34).

We demonstrate endogenous gene activation in which AvrXa10, DOF1, and DREB2 consistently produce the greatest activation across all genes tested. It is interesting to note that the relative strength of ADs did vary both across species and across target loci. For example, our results in *Z. mays* protoplasts show the AvrXa10 domain having reduced activity relative to *A. thaliana* dual luciferase assays. Further, the AvrXa10 domain also shows the largest range of activation among the three strongest ADs across target genes in *A. thaliana* protoplasts. Given that AvrXa10 is derived from a plant pathogen, perhaps there are varying degrees of host response to the expression of this domain across plant species and/or target gene.

We also observed what we will call 'jackpot activation' in which the maximum observed fold change for a given activation domain at a target gene vastly outperforms its median activation value across all target genes. While this can be attributed to lower numbers of biological replicates, we observed jackpot activation across all three strong activation domains in this study. Given that transcription is a dynamic and noisy process, it is plausible that jackpot activation is a natural consequence of recruiting the AvrXa10, DOF1, and/or DREB2 domains to a core gene promoter. Many activation domains contain acidic residues, and it has been suggested that these acidic residues result in phase separation at a gene promoter to form a molecular condensate. It is through this process that the mediator complex is recruited, resulting in gene activation through stabilization of the transcription preinitiation complex (21). It is plausible that jackpot activation could be a consequence of variable co-factor recruitment to the core promoter that is being targeted by our dCas9 PTAs. Future investigation into what co-factors are recruited by AvrXa10, DOF1, and DREB2 could provide further insight into core transcriptional machinery that is necessary and/or sufficient to drive gene activation.

We also demonstrate the ability of AvrXa10, DOF1, and DREB2 plant-derived activation domains to drive gene activation of the *FT* locus in transgenic plants. T1 plants showed an early flowering phenotype, and this phenotype was maintained in the T2 generation where early flowering was quantified. This work illustrates the feasibility of moving from a protoplast-based discovery platform to validating the tools in stable transgenic lines.

Finally, we showcase the ability of the DREB2 activation domain to function in a potential gene enhancer discovery pipeline. The FT locus contains three known enhancers of gene expression Block B, C, and E. These enhancers are located vast distances from the core promoter, and fail to produce any observable activation when targeted with dCas9 PTAs. While we do observe slight boosts in FT activation when co-targeting both the core promoter and an intervening region between enhancers, these boosts were rarely statistically significant and never as dramatic as the overexpression seen when co-targeting an enhancer. Noisy gene activation at distances outside of the core promoter (50-400bp upstream TSS) has been reported before (34), and may be a natural consequence of transcription pre-initiation dynamics.

Determining the location and activity of enhancer sequences can be challenging. Luciferase assays in protoplasts have traditionally been used to confirm enhancer activity by cloning an enhancer upstream of the Firefly luciferase reporter and checking for activity (35). PTA-mediated enhancer activation retains the original genomic architecture and chromatin state of the enhancer-promoter pair, while traditional luciferase assays deploy enhancer sequences in non-native genomic architectures on plasmids which lack nucleosomes. Given our understanding of enhancers functioning through a DNA looping mechanism, retaining this genetic architecture may be important.

Once a putative enhancer is identified it can also be difficult to determine the promoter, or promoters, it regulates. Recent studies have utilized methods such as STARR-seq to confirm the transcriptional regulatory capacity of distal accessible chromatin elements in protoplasts (36). We anticipate PTA-mediated enhancer activation can be similarly used for testing distal accessible chromatin activity in plants. One could design a series of sgRNAs to target intergenic regions in tandem with a sgRNA targeting the core promoter for a gene of interest. Conversely, an accessible chromatin region of interest can be targeted along with the core promoters of nearby genes to identify which transcripts are significantly activated by the enhancer. When conducted in a high-throughput transient platform such as protoplasts, this pipeline could reveal novel enhancer-promoter interactions. This type of enhancer mapping can further expand the definition of gene by including enhancer regions in addition to the core promoter, untranslated regions, exons, introns, and terminator comprising a transcriptional block.

This study has limitations that should be addressed with future work. The PTAs we describe here work great in research and development environments, but they have components derived from human or plant pathogens that would trigger

DRAFT

special regulatory approval considerations if used in transgenic plants intended for field release. Replacing the VP64 domain with plant-evolved ADs mitigates that in part, however the dCas9, GB1, and AvrXa10 components are similarly derived from human or plant pathogens.

In addition, it has been noted that scFv proteins have difficulty forming disulfide bonds within the reducing environment of the cytosol, leading to poor solubility (37). For these reasons, replacing GB1 with alternate solubility tags could be addressed to further improve the performance of these PTAs in plants. For example, a naturally occurring solubility tag from spider silk shows promise as an alternative to the *Streptococcal* GB1 sequence (38). Improving the solubility of the PTA complex without compromising its transcription activation potency should be of significant interest moving forward. With this work we showcase potent PTAs in plants capable of driving strong expression of target genes. We present three transcription activation domains from plant backgrounds that outperform VP64 in both reporter gene assays and targeting endogenous loci. We further showcase the portability of both DREB2 and DOF1 in diverse plants. They provide stronger expression than VP64 for target genes in *A. thaliana*, *S. viridis*, and *Z. mays*. Finally, we show that the DREB2 domain is equally as potent when fused to an alternative TALE programmable DNA binding domain. Given these results, we anticipate that these domains will be used by plant scientists to effectively activate transcription of target genes within a model crop.

ACKNOWLEDGEMENTS

MHZ and DV are supported by Department of Energy Grant DE-SC0018277. MHZ, JACM, AS, and MJS are supported by the USDA grant 2018-33522-28747 and are supported by the Advanced Plant Technologies program, DARPA Award HR001118C0146. MHZ and AS are supported by an NIH NIGMS Biotechnology Training grant NIHT32GM008347.

AUTHOR CONTRIBUTIONS

MHZ, DV, and MJS conceived this study. MHZ designed and built constructs, planned experiments, collected and analyzed data. ACM, AS, and MHZ performed protoplast isolation, transformation, RNA extraction, and RT-qPCR analysis. ACM built constructs and carried out floral dip. ACM provided VP64 transgenic line for FT analysis. MHZ, MJS, and DV wrote the manuscript.

COMPETING FINANCIAL INTERESTS

MHZ, MJS, and DV are listed as inventors on UMN-owned patents related to this work.

Bibliography

1. Anna V. Klepikova, Artem S. Kasianov, Evgeny S. Gerasimov, Maria D. Logacheva, and Aleksey A. Penin. A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *The Plant Journal*, 88(6):1058–1070, dec 2016. ISSN 1365-313X. doi: 10.1111/TPJ.13312.
2. Steffen Knauer, Marie Javelle, Lin Li, Xianran Li, Xiaoli Ma, Kokulapalan Wimalanathan, Sunita Kumari, Robyn Johnston, Samuel Leiboff, Robert Meeley, Patrick S. Schnable, Doreen Ware, Carolyn Lawrence-Dill, Jianming Yu, Gary J. Muehlbauer, Michael J. Scanlon, and Marja C.P. Timmermans. A high-resolution gene expression atlas links dedicated meristem genes to key architectural traits. *Genome Research*, 29(12):1962–1973, dec 2019. ISSN 1088-9051. doi: 10.1101/GR.250878.119.
3. Yunlong Pang, Chunxia Liu, Danfeng Wang, Paul St. Amand, Amy Bernardo, Wenhui Li, Fang He, Linzhi Li, Liming Wang, Xiufang Yuan, Lei Dong, Yu Su, Huirui Zhang, Meng Zhao, Yunlong Liang, Hongze Jia, Xitong Shen, Yue Lu, Hongming Jiang, Yuye Wu, Anfei Li, Honggang Wang, Lingrang Kong, Guihua Bai, and Shubing Liu. High-Resolution Genome-wide Association Study Identifies Genomic Regions and Candidate Genes for Important Agronomic Traits in Wheat. *Molecular Plant*, 13(9):1311–1327, sep 2020. ISSN 1674-2052. doi: 10.1016/J.MOLP.2020.07.008.
4. Fei He, Wei Wang, William B. Rutter, Katherine W. Jordan, Jie Ren, Ellie Taagen, Noah DeWitt, Deepmala Sehgal, Sivakumar Sukumaran, Susanne Dreisigacker, Matthew Reynolds, Jyotirmoy Halder, Sunish Kumar Sehgal, Shuyu Liu, Jianli Chen, Allan Fritz, Jason Cook, Gina Brown-Guedira, Mike Pumphrey, Arron Carter, Mark Sorrells, Jorge Dubcovsky, Matthew J. Hayden, Alina Akhunova, Peter L. Morrell, Les Szabo, Matthew Rouse, and Eduard Akhunov. Genomic variants affecting homoeologous gene expression dosage contribute to agronomic trait variation in allopolyploid wheat. *Nature Communications 2022 13:1*, 13(1):1–15, feb 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-28453-y.
5. J. Armando Casas-Mollano, Matthew H. Zinselmeier, Samuel E. Erickson, and Michael J. Smanski. CRISPR-Cas Activators for Engineering Gene Expression in Higher Eukaryotes. *CRISPR Journal*, 3(5):350–364, oct 2020. ISSN 25731602. doi: 10.1089/CRISPR.2020.0064/ASSET/IMAGES/LARGE/CRISPR.2020.0064_FIGURE2.JPEG.
6. Pablo Perez-Pinera, D. Dewran Kocak, Christopher M. Vockley, Andrew F. Adler, Ami M. Kabadi, Lauren R. Polstein, Pratiksha I. Thakore, Katherine A. Glass, David G. Ousterout, Kam W. Leong, Farshid Guilak, Gregory E. Crawford, Timothy E. Reddy, and Charles A. Gersbach. RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nature Methods 2013 10:10*, 10(10):973–976, jul 2013. ISSN 1548-7105. doi: 10.1038/nmeth.2600.
7. Morgan L. Maeder, Samantha J. Linder, Vincent M. Cascio, Yanfang Fu, Quan H. Ho, and J. Keith Joung. CRISPR RNA–guided activation of endogenous human genes. *Nature Methods 2013 10:10*, 10(10):977–979, jul 2013. ISSN 1548-7105. doi: 10.1038/nmeth.2598.
8. Anna M. Chiarella, Kyle V. Butler, Berkley E. Gryder, Dongbo Lu, Tiffany A. Wang, Xufen Yu, Silvia Pomella, Javed Khan, Jian Jin, and Nathaniel A. Hathaway. Dose-dependent activation of gene expression is achieved using CRISPR and small molecules that recruit endogenous chromatin machinery. *Nature biotechnology*, 38(1):50, jan 2020. ISSN 15461696. doi: 10.1038/S41587-019-0296-7.
9. Haibo Zhou, Junlai Liu, Changyang Zhou, Ni Gao, Zhiping Rao, He Li, Xinde Hu, Changlin Li, Xuan Yao, Xiaowen Shen, Yidi Sun, Yu Wei, Fei Liu, Wenqin Ying, Junming Zhang, Cheng Tang, Xu Zhang, Huatai Xu, Linyu Shi, Leping Cheng, Pengyu Huang, and Hui Yang. In vivo simultaneous transcriptional activation of multiple genes in the brain using CRISPR–dCas9-activator transgenic mice. *Nature Neuroscience 2018 21:3*, 21(3):440–446, jan 2018. ISSN 1546-1726. doi: 10.1038/s41593-017-0060-6.
10. Ambuj Upadhyay, Nathan R. Feltman, Adam Sychla, Anna Janzen, Siba R. Das, Maciej Maselko, and Michael Smanski. Genetically engineered insects with sex-selection and genetic incompatibility enable population suppression. *eLife*, 11, feb 2022. ISSN 2050084X. doi: 10.7554/ELIFE.71230.
11. Ivan Sadowski, Jun Ma, Steve Triezenberg, and Mark Ptashne. GAL4-VP16 is an unusually potent transcriptional activator. *Nature 1988 335:6190*, 335(6190):563–564, 1988. ISSN 1476-4687. doi: 10.1038/335563a0.
12. Roger R. Beerli, David J. Segal, Birgit Dreier, and Carlos F. Barbas. Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25):14628–14633, dec 1998. ISSN 00278424. doi: 10.1073/PNAS.95.25.14628.
13. Levi G. Lowder, Jianping Zhou, Yingxiao Zhang, Aimee Malzahn, Zhaohui Zhong, Tzung Fu Hsieh, Daniel F. Voytas, Yong Zhang, and Yiping Qi. Robust Transcriptional Activation in Plants Using Multiplexed CRISPR-Act2.0 and mTALE-Act Systems. *Molecular Plant*, 11(2):245–256, feb 2018. ISSN 1674-2052. doi: 10.1016/J.MOLP.2017.11.010.
14. Alejandro Chavez, Jonathan Scheiman, Suhani Vora, Benjamin W Pruitt, Marcelle Tuttle, Eswar P R Iyer, Shuailiang Lin, Samira Kiani, Christopher D Guzman, Daniel J Wiegand, Dmitry Ter-Ovanesyan, Jonathan L Braff, Noah Davidsohn, Benjamin E Housden, Norbert Perrimon, Ron Weiss, John Aach, James J Collins, and George M Church. Highly efficient Cas9-mediated transcriptional programming. *Nature Methods 2015 12:4*, 12(4):326–328, mar 2015. ISSN 1548-7105. doi: 10.1038/nmeth.3312.
15. Zhenxiang Li, Dandan Zhang, Xiangyu Xiong, Bingyu Yan, Wei Xie, Jen Sheen, and Jian Feng Li. A potent Cas9-derived gene activator for plant and mammalian cells. *Nature Plants 2017 3:12*, 3(12):930–936, nov 2017. ISSN 2055-0278. doi: 10.1038/s41477-017-0046-0.
16. Marvin E. Tanenbaum, Luke A. Gilbert, Lei S. Qi, Jonathan S. Weissman, and Ronald D. Vale. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell*, 159(3):635–646, oct 2014. ISSN 0092-8674. doi: 10.1016/J.CELL.2014.09.039.
17. Ashot Papikian, Wanlu Liu, Javier Gallego-Bartolomé, and Steven E. Jacobsen. Site-specific manipulation of Arabidopsis loci using CRISPR-Cas9 SunTag systems. *Nature Communications 2019 10:1*, 10(1):1–11, feb 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-08736-7.
18. Sang Dong Yoo, Young Hee Cho, and Jen Sheen. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols 2007 2:7*, 2(7):1565–1572, jun 2007. ISSN 1750-2799. doi: 10.1038/nprot.2007.199.
19. Adam Sychla, Juan Armando Casas-Mollano, Matthew H. Zinselmeier, and Michael Smanski. Characterization of Programmable Transcription Activators in the Model Monocot Setaria viridis Via Protoplast Transfection. In Kan Wang and Zhang Feng, editors, *Protoplast Technology, Methods in Molecular Biology*, volume 2464, pages 223–244. Humana, New York, NY, 2022. doi: 10.1007/978-1-0716-2164-6_16.
20. Tomáš Čermák, Shaun J. Curtin, Javier Gil-Humanes, Radim Čegan, Thomas J.Y. Kono, Eva Konečná, Joseph J. Belanto, Colby G. Starker, Jade W. Mathre, Rebecca L. Greenstein, and Daniel F. Voytas. A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. *The Plant Cell*, 29(6):1196–1217, jul 2017. ISSN 1040-4651. doi: 10.1105/TPC.16.00922.
21. Ann Boija, Isaac A. Klein, Benjamin R. Sabari, Alessandra Dall'Agnese, Eliot L. Coffey, Alicia V. Zamudio, Charles H. Li, Krishna Shrinivas, John C. Manteiga, Nancy M. Hannett, Brian J. Abraham, Lena K. Afeyan, Yang E. Guo, Jenna K. Rimel, Charli B. Fant, Jurian Schuijers, Tong Ihn Lee, Dylan J. Taatjes, and Richard A. Young. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*, 175(7):1842–1855.e16, dec 2018. ISSN 0092-8674. doi: 10.1016/J.CELL.2018.10.042.
22. Bruce A Sherf, Shauna L Navarro, Rita R Hannah, and Keith V Wood. Dual-Luciferase TM Reporter Assay: An Advanced Co-Reporter Technology Integrating Firefly and Renilla Luciferase Assays. *Promega Notes Magazine Number*, 57:2, 1996.
23. Kenneth R. Luehrsen, Jeffrey R. de Wet, and Virginia Walbot. [35] Transient expression analysis in plants using firefly luciferase reporter gene. *Methods in Enzymology*, 216(C):397–414, jan 1992. ISSN 0076-6879. doi: 10.1016/0076-6879(92)16037-K.

DRAFT

24. Philip N. Benfey and Nam Hai Chua. The Cauliflower Mosaic Virus 35S Promoter: Combinatorial Regulation of Transcription in Plants. *Science*, 250(4983):959–966, nov 1990. ISSN 00368075. doi: 10.1126/SCIENCE.250.4983.959.
25. Joanne E. Lee, Manuela Neumann, Daniel Iglesias Duro, and Markus Schmid. CRISPR-based tools for targeted transcriptional and epigenetic regulation in plants. *PLOS ONE*, 14 (9):e0222778, sep 2019. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0222778.
26. S. Boersma, Deepak Khuperkar, Bram M.P. Verhagen, S. Sonneveld, Jonathan B. Grimm, Luke D. Lavis, and Marvin E. Tanenbaum. Multi-Color Single-Molecule Imaging Uncovers Extensive Heterogeneity in mRNA Decoding. *Cell*, 178(2):458–472.e19, jul 2019. ISSN 0092-8674. doi: 10.1016/J.CELL.2019.05.001.
27. Steven J. Clough and Andrew F. Bent. Floral dip: a simplified method for Agrobacterium -mediated transformation of Arabidopsis thaliana. *The Plant Journal*, 16(6):735–743, dec 1998. ISSN 1365-313X. doi: 10.1046/J.1365-313X.1998.00343.X.
28. Takashi L. Shimada, Tomoo Shimada, and Ikuko Hara-Nishimura. A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana. *The Plant Journal*, 61(3):519–528, feb 2010. ISSN 1365-313X. doi: 10.1111/J.1365-313X.2009.04060.X.
29. Juan Armando Casas-Mollano, Matthew H. Zinselmeier, and Michael J. Smanski. Efficient gene activation in plants by the MoonTag programmable transcriptional activator. *Manuscript in preparation*, 2022.
30. Shinobu Takada and Koji Goto. TERMINAL FLOWER2, an Arabidopsis Homolog of HETEROCHROMATIN PROTEIN1, Counteracts the Activation of FLOWERING LOCUS T by CONSTANS in the Vascular Tissues of Leaves to Regulate Flowering Time. *The Plant Cell*, 15(12):2856–2865, dec 2003. ISSN 1040-4651. doi: 10.1105/TPC.016345.
31. Martín Krzymuski, Fernando Andrés, Juan I. Cagnola, Seonghoe Jang, Marcelo J. Yanovsky, George Coupland, and Jorge J. Casal. The dynamics of FLOWERING LOCUS T expression encodes long-day information. *The Plant Journal*, 83(6):952–961, sep 2015. ISSN 1365-313X. doi: 10.1111/TPJ.12938.
32. Y. Esther Tak, Joy E. Horng, Nicholas T. Perry, Hayley T. Schultz, Sowmya Iyer, Qiuming Yao, Luli S. Zou, Martin J. Aryee, Luca Pinello, and J. Keith Joung. Augmenting and directing long-range CRISPR-mediated activation in human cells. *Nature Methods 2021 18:9*, 18(9): 1075–1081, aug 2021. ISSN 1548-7105. doi: 10.1038/s41592-021-01224-1.
33. Johan Zicola, Liangyu Liu, Petra Tänzler, and Franziska Turck. Targeted DNA methylation represses two enhancers of FLOWERING LOCUS T in Arabidopsis thaliana. *Nature Plants 2019 5:3*, 5(3):300–307, mar 2019. ISSN 2055-0278. doi: 10.1038/s41477-019-0375-2.
34. Luke A. Gilbert, Max A. Horlbeck, Britt Adamson, Jacqueline E. Villalta, Yuwen Chen, Evan H. Whitehead, Carla Guimaraes, Barbara Panning, Hidde L. Ploegh, Michael C. Bassik, Lei S. Qi, Martin Kampmann, and Jonathan S. Weissman. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, 159(3):647–661, oct 2014. ISSN 0092-8674. doi: 10.1016/J.CELL.2014.09.029.
35. Anthony Studer, Qiong Zhao, Jeffrey Ross-Ibarra, and John Doebley. Identification of a functional transposon insertion in the maize domestication gene tb1. *Nature Genetics 2011 43:11*, 43(11):1160–1163, sep 2011. ISSN 1546-1718. doi: 10.1038/ng.942.
36. William A. Ricci, Zefu Lu, Lexiang Ji, Alexandre P. Marand, Christina L. Ethridge, Nathalie G. Murphy, Jaclyn M. Noshay, Mary Galli, María Katherine Mejía-Guerra, Maria Colomé-Tatché, Frank Johannes, M. Jordan Rowley, Victor G. Corces, Jixian Zhai, Michael J. Scanlon, Edward S. Buckler, Andrea Gallavotti, Nathan M. Springer, Robert J. Schmitz, and Xiaoyu Zhang. Widespread long-range cis-regulatory elements in the maize genome. *Nature Plants 2019 5:12*, 5(12):1237–1249, nov 2019. ISSN 2055-0278. doi: 10.1038/s41477-019-0547-0.
37. Adam C. Fisher and Matthew P. DeLisa. Efficient isolation of soluble intracellular single-chain antibodies using the twin-arginine translocation machinery. *Journal of molecular biology*, 385(1):299, jan 2009. ISSN 00222836. doi: 10.1016/J.JMB.2008.10.051.
38. Nina Kronqvist, Medoune Sarr, Anton Lindqvist, Kerstin Nordling, Martins Otikovs, Luca Venturi, Barbara Pioselli, Pasi Purhonen, Michael Landreh, Henrik Biverstål, Zigmantas Toleikis, Lisa Sjöberg, Carol V. Robinson, Nicola Pelizzi, Hans Jörnvall, Hans Hebert, Kristaps Jaudzems, Tore Curstedt, Anna Rising, and Jan Johansson. Efficient protein production inspired by how spiders make silk. *Nature Communications 2017 8:1*, 8(1):1–15, may 2017. ISSN 2041-1723. doi: 10.1038/ncomms15504.
39. Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, oct 1990. ISSN 0022-2836. doi: 10.1016/S0022-2836(05)80360-2.
40. Jianmei Cao, Dongmei Yao, Fan Lin, and Mingyi Jiang. PEG-mediated transient gene expression and silencing system in maize mesophyll protoplasts: A valuable tool for signal transduction study in maize. *Acta Physiologiae Plantarum*, 36(5):1271–1281, mar 2014. ISSN 01375881. doi: 10.1007/S11738-014-1508-X/FIGURES/6.

DRAFT

Supplementary Note 1: Materials and Methods

Plasmid construction. Putative activation domains were selected from literature analysis for the potential to recruit transcription initiation machinery in plant cells. DNA binding domains from transcription factors were identified and removed on the basis of sequence conservation to experimentally determined DNA binding domain motifs. Patches of acidic/aromatic residues were identified as core motifs for a prospective activation domain, and a given sequence was selected to encompass as many of these patches as possible without including the DNA binding domain. Synthetic DNA fragments were then designed and codon optimized based on average codon usage in *A. thaliana* and *Oryza sativa*. Fragments were synthesized by Genscript with BsmBI cloning sites flanking the prospective activation domain. Fragments were cloned into an scFv destination vector with BsaI and BsmBI cloning sites generating compatible overhangs. The assembled plasmids generate an in-frame C terminal fusion of the activation domain to the scFv. The scFv is driven by the CaMV 35S promoter, and also contains a C terminal fusion of the GB1 solubility tag. The scFv-AD is recruited to targets by the dCas9-GCN4. This vector was generated by fusing 24 copies of the GCN4 epitope tag to the C-terminus of *A. thaliana* codon optimized dCas9. Each GCN4 repeat was separated by short linker GS linker sequences. The dCas9 is expressed under the AtUbi10 promoter in dicots, the CMYLCV promoter in *S. viridis*, or the ZmUbi10 promoter in Zea mays. Guide RNAs were designed and expressed from U6 promoters, either the *A. thaliana* U6 promoter for dicots or the *Oryza sativa* U6 promoter for monocots. A complete list of plasmids used in this study can be found in Supplemental Table 1.

PTA protein sequence similarity. BLAST (ver. 2.8.1+)(39) was used to perform an all-by-all comparison of the proteins from which the activators were derived. Full length amino acid sequences were used as both query and subject, and no e-value threshold was set. Expectation values were negative log transformed for plotting.

Protoplast isolation and transformation. *S. viridis* and *A. thaliana* protoplasts were isolated and transformed as stated in previous protocols(18, 19). ME034 *S. viridis* plants are grown in a growth chamber set to 31/22 degrees C with 12h diurnal cycle. Mesophyll cells can be isolated from leaf tissue 14-21 days post germination. Col-0 *A. thaliana* plants are grown in a growth chamber set to 22C with 16h light cycle. Mesophyll cells can be isolated from leaf tissue 14-21 days post germination. Zea mays protoplasts were isolated and transformed as stated in previous protocols. (40) Zea mays plants were grown in a growth chamber at 25 degrees C under 16h light cycle. Prior to isolation, plants were 'greened' by first placing the seedlings in the dark for 5 days post-germination followed by 2 days of exposure to light prior to isolation. Plasmids to be transformed in all systems are prepared according to the manufacturer's protocol (Qiagen 12945) to ensure transformation grade endotoxin free DNA.

Dual luciferase assay. A dual luciferase assay was designed with the Firefly and Renilla luciferase reporter genes from Promega (22). A Renilla luciferase was cloned upstream of a Firefly luciferase on a single plasmid. The Renilla CDS was under the control of a constitutive promoter, either the CaMV 35S or weaker *Nos* promoter. Upstream of the Firefly CDS was a minimal CaMV 35S promoter, containing 60 bp upstream of the TSS. Six copies of the Lac Operator were cloned upstream of the minimal promoter driving Firefly. An sgRNA targeting the LacO region results in targeting of the dCas9 activation to the Firefly promoter. We drove dCas9-24xGCN4 with a constitutive promoter (CMYLCV, AtUbi10, or ZmUbi1), the LacO sgRNA driven by *Oryza sativa* U6 promoter, and the scFv with a constitutive 35S promoter. We co-transformed *S. viridis*, *A. thaliana*, or *Zea mays* protoplasts with the dCas9-24xGCN4, *Lac*O sgRNA, scFv-AD, and dual luciferase reporter plasmids. Following 24h incubation at 25C, we lysed the cells in 1X Passive Lysis Buffer from Promega. We then quantified Firefly and Renilla luciferase luminescence using the GloMax Explorer plate reader (Promega GM3500) equipped with dual injectors with the Dual Luciferase Assay kit (Promega E1960). Firefly luciferase substrate was injected first and luminescence was quantified, followed by injection of the Renilla luciferase substrate and subsequent quantification. We normalized for delivery first by dividing Firefly by Renilla activity unless otherwise described, followed by Fold Change calculation by dividing a given AD by the negative control or VP64. A complete list of plasmids used in this study can be found in Supplemental Table 2.

RNA isolation and quantification. We isolate RNA from either protoplasts or leaf tissue according the the TRIzol manufacturer's protocol (Thermo 15596026). For protoplasts we spun the cells down at 500g for 2 minutes, followed by removal of W5 buffer before re-suspending cells in 1mL of TRIzol. RNA was collected 24h post-transformation for protoplast samples. For plant tissue samples, we first snap froze the sample in liquid nitrogen followed by shaking in a paint shaker apparatus with metal beads to homogenize frozen tissue. We then add 1mL of TRIzol to the homogenized tissue. We then follow the TRIzol protocol according to manufacturer specifications. We then treat the samples with Turbo DNA Free Kit (Invitrogen AM1907) to remove any plasmid and/or genomic DNA from the RNA sample. To quantify a transcript we then perform RT-qPCR using gene-specific primer pairs. A list of primers used for quantification can be found in Supplemental Table 2. We follow a RT-qPCR cycling protocol as defined by the NEB Luna One-Step RT-qPCR kit manufacturer's protocol (NEB E3005L). Primers are designed to have Tm values of 60C and yield amplicon lengths between 75-175bp. The primers typically span an intron, such that the shorter PCR product corresponds to spliced mRNA. We quantify gene expression for relative comparison between treatments using the delta delta Ct method. All primers can be found in Supplemental Table 2.

DRAFT

Plant transformation and genotyping. To generate transgenic *A. thaliana* plants we performed floral dip protocols as previously described (27). Binary vectors were constructed containing an antibiotic resistance gene along with the pFAST Oleosin-RFP transgene (28). *A. thaliana* ecotype Columbia-0 was grown to maturity and floral dipped with Agrobacterium strain GV3101 transformed with the described binary vectors. The antibiotic selection used in this study was either Kanamycin or Bialaphos resistance. T0 seeds were identified by the pFAST system in which RFP-positive seedlings contain the transgene. We placed RFP+ seedlings directly onto soil, or onto selective media, to grow T1 plants. T1 plants were grown to maturity and allowed to set seed. All molecular characterization was performed on T2 lines, along with phenotyping. To quantify FT overexpression phenotypes in T2 plants, we planted 18 RFP positive T2 plants from each T1 parent. We chose time point 0 to be the date at which the first set of true leaves emerge from a seedling. We then count the number of days elapsed until bolting is observed. We also count the number of rosette leaves one day after bolting was observed. RT-qPCR was performed to quantify target gene expression, using primers indicated in Supplemental Table 2.

DRAFT

Supplementary Note 2: Supplementary Table S1. Activation domains

Name	Domain Source	AA range	AD Seq	DOI	Uniprot
DOF1	Plant	175-238	QPGTEDAEAV ALGLGLSDFP SAGKAVLDDE DSFVWPAASF DMGACWAGAG FADPDPACIF LNLP	10.1093/pcp/pce105	http://www.uniprot.org/uniprot/P38564
DREB1	Plant	108-216	GRSACLNFAD SAWRLRIPES TCAKDIQKAA AEAALAFQDE MCDATTDHGF DMEETLVEAI YTAEQSE-NAF YMHDEAMFEM PSLLANMAEG MLLPLPSVQW NHNHEVDGDD DDVSLWSY	10.1111/pbi.12057	http://www.uniprot.org/uniprot/Q9M0L0
DREB2	Plant	254-335	SSDMFDVDEL LRDLNGDDVF AGLNQDRYPG NSVANGSYRP ESQQSGFDPL QSLNYGIPPF QLEGKDGNGF FDDLSYLDLE N	10.1111/pbi.12057	http://www.uniprot.org/uniprot/O82132
IPN2	Plant	179-358	AATNLKGIGP QTIPDMGIMK EFGSPLGFSF QDLDLYGGGG GDQLELQQNM EKPPLDGFMP MNHENL-CLGK KRPNPYSGNN GKSPLMWSDD LRLQDLGSCL QDDPFKGDHH HQIQIAPPSL DRGTEMDPMS EIYDSKPEEK KFDASMKLER PSPRRAPLGE RMSPMITTGT MAQGRSSPFG	10.1111/nph.12593	https://www.uniprot.org/uniprot/E5L8F7
PTI4	Plant	170-234	EPEPVRVTAK RRASPEPASS SGNGSMKRRR KAVQKCDGEM ASRSSVMQVG CQIEQLTGVH QLLVI	10.1111/pbi.12057	http://www.uniprot.org/uniprot/O04680
AtARF8	Plant	350-702	VSLWEIEPLT TFPMYPSLFP LRLKRPWHAG TSSLPDGRGD LGSGLTWLRG GGGEQQGLLP LNYPSVGLFP WMQQRLDLSQ MGTDNNQQYQ AMLAAGLQNI GGGDPLRQQF VQLQEPHHQY LQQSASHNSD LMLQQQQQQQ ASRHLMHAQT QIMSENLPQQ NMRQEVSNQP AGQQQQLQQP DQNAYLNAFK MQNGHLQQWQ QQSEMPSPSF MKSDFTDSSN KFATTASPAS GDGNLLNFSI TGQSVLPEQL TTEGWSPKAS NTFSEPLSLP QAYPGKSLAL EPGNPQNPSL FGVDPDSGLF LPSTVPRFAS SSGDAEASPM SLTDSGFQNS LYSCMQDTTH ELLHGAGQIN S	10.1105/tpc.008417	https://www.uniprot.org/uniprot/Q9FGV1
SNAC1	Plant	172-314	WEKMQQGKEV KEEASDMVTS QSHSHTHSWG ETRTPESEIV DNDPFPELDS FPAFQPAPPP ATAMMVP-KKE SMDDATAAAA AAATIPRNNS SLFVDLSYDD IQGMYSGLDM LPPGDDFYSS LFASPRVKGT TPRA-GAGMGM VPF	10.1073/pnas.0604882103	http://www.uniprot.org/uniprot/A2XNB9
Opaque-2	Plant	41-227	IDVAAAGHGD GDMMDQQHAT EWTFERLLEE EALTTSTPPP VVVVPNSCCS GALNADRPPV MEEAVTMAPA AVSSAVVGDP MEYNAILRRK LEEDLEAFKM WRADSSVVTS DQRSQGSNNH TGGSSIRNNP VQNKLMNGED PINNNHAQTA GLGVRLATSS SSRDPSPSDE DMDGEVEILG FKM	PMC146487	http://www.uniprot.org/uniprot/P12959
AvrXa10	Bacterial		TVMWEQDAAP FAGAADDFPA FNEEELAWLM ELLPQSGSVG GTI	10.1094/MPMI.1998.11.8.824	https://www.uniprot.org/uniprot/Q56830
AL2/C2 /TrAP	Viral	83-129	PLHQHQDIPL TNQVQPQPEE SIGSPQGISQ LPSMDDIDDS FWENLFK	10.1006/viro.1999.9925	http://www.uniprot.org/uniprot/P03562
AtARR2	Plant	279-664	QHQGNMNHSF MTGQDQSFGP LSSLNGFDLQ SLAVTGQLPP QSLAQLQAAG LGRPTLAKPG MSVS-PLVDQR SIFNFENPKI RFGDGHGQTM NNGNLLHGVP TGSHMRLRPG QNVQSSGMML PVADQLPRGG PSMLPSLGQQ PILSSSVSRR SDLTGALAVR NSIPETNSRV LPTTHSVFNN FPADLPRSSF PLASAPGISV PVSVSYQEEV NSSDAKGGSS AATAGFGNPS YDIFNDFPQH QQHNKNISNK LNDWDLRNMG LVFSSNQ-DAA TATATAAFST SEAYSSSSTQ RKRRETDATV VGEHGQNLQS PSRNLYHLNH VFMDGGSVRV KSER-VAETVT CPPANTLFHE QYNQEDLMSA FLKQEGIPSV DNEFEFDGYS IDNIQV	10.1111/j.1365-313X.2000.00909.x	https://www.uniprot.org/uniprot/Q9ZWJ9
RF2a	Plant	56-108	EILSLPEDLD LCAAGGGDGP SLSDENDEEL FSMFLDVEKL NSTCGASSEA EAE	10.1074/jbc.M304862200	http://www.uniprot.org/uniprot/Q69IL4
RF2a	Plant	283-357	GGMMMNFGGM PHQFGGNQQM FQNNQAMQSM LAAHQLQQLQ LHPQAQQQQV LHPQHQQQQP LH-PLQAQQLQ QAARDLKMKS PMGGQSQWGD GKSGSSGN	10.1074/jbc.M304862200	http://www.uniprot.org/uniprot/Q69IL4
OsCBT	Plant	542-755	PVTESLLELV LRNRLQEWLV EMVMEGHKST GRDDLGQGAI HLCSFLGYTW AIRLFSLSGF SLD-FRDSSGW TALHWAAYHG RERMVATLLS AGANPSLVTD PTPESPAGLT AADLAARQGY DGLAAY-LAEK GLTAHFEAMS LSKDTEQSPS KTRLTKLQSE KFEHLSEQEL CLKESLAAYR NAADAASNIQ AAL-RERTLKL QTKAIQLANP EIEASEIV	10.1074/jbc.M504616200	https://www.uniprot.org/uniprot/Q7XHR2
AP1	Plant	193-256	HNMPPPLPPQ QHQIQHPYML SHQPSPFLNM GGLYQEDDPM AMRRNDLELT LEPVYNCNLG CFAA	10.1023/A:1006273127067	https://www.uniprot.org/uniprot/P35631
MYC2	Plant	148-188	GGVAPSDDAV DEEVTDTEWF FLVSMTQSFA CGAGLAGKAF A	10.1016/j.celrep.2017.04.057	https://www.uniprot.org/uniprot/Q39204
ZmVP1	Plant	1-120	MEASSGSSPP HSQENPPEHG GDMGGAPAEE IGGEAADDFM FAEDTFPSLP DFPCLSSPSS STFSSNSSSN SSSAYTNTAG RAGGEPSEPA SAGEGFDALD DIDQLLDFAS LSMPWDSEPF	10.1016/0092-86749190436-3	https://www.uniprot.org/uniprot/P26307

DRAFT

Supplementary Note 3: Supplementary Table S1. Activation domains continued

Name	Domain Source	AA range	AD Seq	DOI	Uniprot
OsGRF1	Plant	221-396	PPPSYYSMDH KEYAYGHATK EVHGEHAFFS DGTEREHHHA AAGHGQWQFK QLGMEPKQST TPLFP-GAGYG HTAASPYAID LSKEDDDEKE RRQQQQQQQQ HCFLLGADLR LEKPAGHDHA AAAQKPLRHF FDEWPHEKNS KGSWMGLEGE TQLSMSIPMA ANDLPITTTS RYHNDE	10.1093/pcp/pch098	https://www.uniprot.org/uniprot/A2XA73
AtHSFA2	Plant	230-341	KEKKSLFGLD VGRKRRLTST PSLGTMEENL LHDQEFDRMK DDMEMLFAAA IDDEANNSMP TKEEQ-CLEAM NVMMRDGNLE AALDVKVEDL VGSPLDWDSQ DLHD	10.1111/j.1365-313X.2004.02111.x	https://www.uniprot.org/uniprot/O80982
AtHSFA6b	Plant	252-391	KEKRKEIEEA ISKKRQRPID QGKRNVEDYG DESGYGNDVA ASSSALIGMS QEYTYGNMSE FEMSELD-KLA MHIQGLGDNS SAREEVLNVE KGNDEEEVED QQQGYHKENN EIYGEGFWED LLNEGQNFDF EGDQENVDVG SSSHTN	10.1111/j.1365-313X.2004.02111.x	https://www.uniprot.org/uniprot/Q9LUH8
EFS	Plant	820-1230	KTTGALLDAD IGKTSATYGT ISSDVTHGEM VVDVTIEDSY STESAWVRCD DCFKWRRIPA SVVG-SIDESS RWICMNNSDK RFADCSKSQE MSNEEINEEL GIGQDEADAY DCDAAKRGKE KEQKSKRLTG KQKACFKAIK TNQFLHRNRK SQTIDEIMVC HCKPSPDGRL GCGEECLNRM LNIECLQGTC PAGDLC-SNQQ FQKRKYVKFE RFQSGKKGYG LRLLEDVREG QFLIEYVGEV LDMQSYETRQ KEYAFKGQKH FYFMTLNGNE VIDAGAKGNL GRFINHSCEP NCRTEKWMVN GEICVGIFSM QDLKKGQELT FDYNYVRVFG AAAKKCYCGS SHCRGYIGGD PLNGDVIIQS DSDEEYPELV ILDDDESGEG ILGATSRTFT DDADEQMPQS FEKVNGYKDL APDNTQTQ	10.1105/tpc.109.070060	https://www.uniprot.org/uniprot/F4I6Z9
AFT1	Plant	Full length	MAATLGRDQY VYMAKLAEQA ERYEEMVQFM EQLVTGATPA EELTVEERNL LSVAYKNVIG SLRAAWRIVS SIEQKEESRK NDEHVSLVKD YRSKVESELS SVCSGILKLL DSHLIPSAGA SESKV-FYLKM KGDYHRYMAE FKSGDERKTA AEDTMLAYKA AQDIAAADMA PTHPIRLGLA LNFSVFYYEI LNSSDKACNM AKQAFEEAIA ELDTLGEESY KDSTLIMQLL RDNLTLWTSD MQEQMDEA	10.1016/S0014-57939801739-6	https://www.uniprot.org/uniprot/P48349
EIN3	Plant	310-620	SLSGGSCSLL MNDCSQYDVE GFEKESHYEV EELKPEKVMN SSNFGMVAKM HDFPVKEEVP AGNSEFMRKR KPNRDLNTIM DRTVFTCENL GCAHSEISRG FLDRNSRDNH QLACPHRDSR LPYGAAPSRF HVNEVKPVVG FPQPRPVNSV AQPIDLTGIV PEDGQKMISE LMSMYDRNVQ SNQTSMVMEN QSVSLLQPTV HNHQEHLQFP GNMVEGSFFE DLNIPNRANN NNSSNNQTFF QGNNNNNNVF KFDTADHNNF EAAHNNNNNS SGNRFQLVFD STPFDMASFD YRDDMSMPGV VGTMDGMQQK QQDVSIWF	10.1016/j.jmb.2005.02.065	https://www.uniprot.org/uniprot/O24606
WRKY50	Plant	1-106	MNDADTNLGS SFSDDTHSVF EFPELDLSDE WMDDDLVSAV SGMNQSYGYQ TSDVAGALFS GSSSCF-SHPE SPSTKTYVAA TATASADNQN KKEKKKIKGR VAFKTR	10.3389/fpls.2018.00930	https://www.uniprot.org/uniprot/Q8VWQ5
ATXR7	Plant	1250-1423	DGADVLKMSQ LKARKKHLRF QQSKIHDWGL VALEPIEAED FVIEYVGELI RSSISEIRER QYEK-MGIGSS YLFRLDDGYV LDATKRGGIA RFINHSCEPN CYTKIISVEG KKKIFIYAKR HIDAGEEISY NYKFPLEDDK IPCNCGAPNV YCFCEQVPWI AKLKRRTWFS RRN	10.1105/tpc.109.070060	https://www.uniprot.org/uniprot/F4K1J4
ASF1a	Plant	Full length	MSAIKITNVA VLHNPAPFVS PFQFEISYEC LNSLKDDLEW KLIYVGSAED ETYDQLLESV LVGPVN-VGNY RFVFQADPPD PSKIQEEDII GVTVLLLTCS YMGQEFLRVG YYVNNDYEDE QLKEEPPTKV LID-KVQRNIL SDKPRVTKFP IDFHPEEEQT AATAAPPEQS DEQQPNVNGE AQVLPDQSVE PKPEES	10.1111/pce.12299	https://www.uniprot.org/uniprot/Q9C9M6
HAG1	Plant	150-550	LESSDGGKDG GSSVVGTGVS GTVGGSSISG LVPKDESVKV LAENFQTSGA YIAREEALKR EEQA-GRLKFV CYSNDSIDEH MMCLIGLKNI FARQLPNMPK EYIVRLLMDR KHKSVMVLRG NLVVGGI-TYR PYHSQKFGEI AFCAITADEQ VKGYGTRLMN HLKQHARDVD GLTHFLTYAD NNAVGYFVKQ GFTKEIYLEK DVWHGFIKDY DGGLLMECKI DPKLPYTDLS SMIRQQRKAI DERIRELSNC QNVYP-KIEFL KNEAGIPRKI IKVEEIRGLR EAGWTPDQWG HTRFKLFNGS ADMVTNQKQL NALMRALLKT MQDHADAWPF KEPVDSRDVP DYYDIIKDPI DLKVIAKRVE SEQYYVTLDM FVADARRMFN NCRTYN-SPDT IYYKCATRLE THFHSKVQAG LQSGAKSQ	10.1093/nar/29.7.1524	https://www.uniprot.org/uniprot/Q9AR19
GT-3A	Plant	110-323	RYKACETTEP DAIRQQFPFY NEIQSIFEAR MQRMLWSEAT EPSTSSKRKH HQFSSDDEEE EVDEP-NQDIN EELLSLVETQ KRETEVITTS TSTNPRKRAK KGKGVASGTK AETAGNTLKD ILEEFMRQTV KMEKEWRDAW EMKEIEREKR EKEWRRRMAE LEEERAATER RWMEREEERR LREEARAQKR DSLI-DALLNR LNRDHNDDHH NQGF	10.1016/S0014-57930400222-4	https://www.uniprot.org/uniprot/Q9SDW0
CVBp12	Virus		LCVDIYKRAF PRSVNKGRSS YAGGSGGVPG ISYNSKVRDY ILWGVTEVIP	10.1105/tpc.112.106476	https://www.uniprot.org/uniprot/P37992
ASF1b	Plant	Full length	MSSINITNVT VLDNPAPFVN PFQFEISYEC LTSLKDDLEW KLIYVGSAED ETYDQVLESV LVGPVN-VGNY RFVLQADSPD PLKIREEDII GVTVLLLTCS YMDQEFIRVG YYVNNDYDDE QLREEPPTKV LID-KVQRNIL TDKPRVTKFP INFHPENEQT LGDGPAPTEP FADSVVNGEA PVFLEQPQKL QEIEQFDDSD VNGEAIALLD QPQNLQET	10.1111/pce.12299	https://www.uniprot.org/uniprot/Q9LS09
TBP1	Plant		MTDQGLEGSN PVDLSKHPSG IVPTLQNIVS TVNLDCKLDL KAIALQARNA EYNPKRFAAV IM-RIREPKTT ALIFASGKMV CTGAKSEDFS KMAARKYARI VQKLGFPAKF KDFKIQNIVG SCDVKFPIRL EGLAYSHAAF SSYEPELFPG LIYRMKVPKI VLLIFVSGKI VITGAKMRDE TYKAFENIYP VLSEFRKIQQ	na	https://www.uniprot.org/uniprot/P28147
BRE1	Plant	700-876	DIQQGSAYAS FLSKKSSRIE DQLRFCTDQF QKLAEDKYQK SVSLENLQKK RADIGNGLEQ ARSR-LEESHS KVEQSRLDYG ALELELEIER FNRRRIEEEM EIAKKKVSRL RSLIEGSSAI QKLRQELSEF KEILKCKACN DRPKEVVITK CYHLFCNPCV QKLTGTRQKK CPTCSASFGP NDIKPIYI	na	https://www.uniprot.org/uniprot/Q8RXD6

DRAFT

Supplementary Note 4: Supplementary Table S1. Activation domains continued

Transcription Factor	Gene ID	AA range	AD sequence	Source	Uniprot
bHLH122	At1g51140	1-300	MESEFQQHHF LLHDHQHQRP RNSGLIRYQS APSSYFSSFG ESIEEFLDRP TSPETERILS GFLQTTDTSD NVDSFLHHTF NSDGTEKKPP EVKTEDEDAE IPVTATATAM EVVVSGDGEI SVNPEVSIGY VASVS-RNKRP REKDDRTPVN NLARHNSSPA GLFSSIDVET AYAAVMKSMG GFGGSNVMST SNTEASSLTP RSKLLPPTSR AMSPISEVDV KPGFSSRLPP RTLSGGFNRS FGNEGSASSK LTALARTQSG GLDQYK-TKDE DSASRRPPLA HHMSLPKSLS DIEQLLSDSI	DAP seq dataset	https://www.uniprot.org/uniprot/Q9C690
LMI1	AT5G03790	140-234	AKQLEQLYDS LRQEYDVVSR EKQMLHDEVK KLRALLRDQG LIKKQISAGT IKVSGEEDTV EISSVVVAHP RTENMNANQI TGGNQVYGQY NNPMLVASSG WPSYP	DAP seq dataset	https://www.uniprot.org/uniprot/Q9LZR0
SDG31	AT3G04380	196-462	CCANCKGNCL SADFPCTCAR ETSGEYAYTK EGLLKEKFLD TCLKMKKEPD SFPKVYCKDC PLERDHDKGT YGKCDGHLIR KFIKECWRKC GCDMQCGNRV VQRGIRCQLQ VYFTQEGKGW GLRTLQDLPK GTFICEYIGE ILTNTELYDR NVRSSSERHT YPVTLDADWG SEKDLKDEEA LCLDAT-ICGN VARFINHRCE DANMIDIPIE IETPDRHYYH IAFFTLRDVK AMDELTWDYM IDFNDKSHPV KAFRCCCGSE SCRDRKI	DAP seq dataset	https://www.uniprot.org/uniprot/Q8W595
VRN1	At3g18990	100-240	EINYHSTGLM DSAHNHFKRA RLFEDLEDED AEVIFPSSVY PSPLPESTVP ANKGYASSAI QTLFTG-PVKA EEPTPTPKIP KKRGRKKKNA DPEEINSSAP RDDDPENRSK FYESASARKR TVTAEERERA IN-AAKTFEPT	DAP seq dataset	https://www.uniprot.org/uniprot/Q8L3W1
ATHB13	AT1G69780	151-294	RQFDTLKAEN DLLQTHNQKL QAEIMGLKNR EQTESINLNK ETEGSCSNRS DNSSDNLRLD ISTAPP-SNDS TLTGGHPPPP QTVGRHFFPP SPATATTTTT TMQFFQNSSS GQSMVKEENS ISNMFCAMDD HS-GFWPWLDQ QQYN	DAP seq dataset	https://www.uniprot.org/uniprot/Q8LC03
ABI5	AT2G36270	1-340	MVTRETKLTS EREVESSMAQ ARHNGGGGGE NHPFTSLGRQ SSIYSLTLDE FQHALCENGK NFGSMN-MDEF LVSIWNAEEN NNNQQQAAAA AGSHSVPANH NGFNNNNNNG GEGGVGVFSG GSRGNEDANN KRGIANESSL PRQGSLTLPA PLCRKTVDEV WSEIHRGGGS GNGGDSNGRS SSSNGQNNAQ NGGE-TAARQP TFGEMTLEDF LVKAGVVREH PTNPKPNPNP NQNQNPSSVI PAAAQQQLYG VFQGTGDPSF PGQAMGVGDP SGYAKRTGGG GYQQAPPVQA GVCYGGGVGF GAGGQQMGMV GPLSPVSSDG LGHGQVDNIG GQYGVDMGGL	DAP seq dataset	https://www.uniprot.org/uniprot/Q9SJN0
LBD2	AT1G06280		FPARKTKEFQ AVHKVFGVSN VQKMVRTVRE EDRTKLSDSL TWEALWRQKD PVLGSYGEYR RICEELKLYK SLVHNQPLIG WDNNQRVFNN NSNNKNGLAM TNSSGSGGFS VNNNGVGVNR EIVNGGYASR NVQGGWENLK HDQRQQCYAV INNGFKQHYL PL	DAP seq dataset	https://www.uniprot.org/uniprot/Q9LNB9
BBX21	AT1G75540	110-331	SSVYKPTSKS SSSSSSNQDF SVPGSSISNP PPLKKPLSAP PQSNKIQPFS KINGGDASVN QWGSTSTISE YLMDTLPGWH VEDFLDSSLP TYGFSKSGDD DGVLPYMEPE DDNNTKRNNN NNNNNNNNTV SLP-SKNLGIW VPQIPQTLPS SYPNQYFSQD NNIQFGMYNK ETSPEVVSFA PIQNMKQQGQ NNKRWYDDGG FTVPQITPPP LSSNKKFRSF W	DAP seq dataset	https://www.uniprot.org/uniprot/Q9LQZ7

DRAFT

Supplementary Note 5: Supplementary Table S2. sgRNA constructs

Name	Link
AtU6 LacO sgRNA	https://benchling.com/s/seq-tOShh49fIMXhHFR2n00n?m=slm-VN5W2LGHQAlKLQ9ZvgQH
AtWUS 4x sgRNA repeat	https://benchling.com/s/seq-flgpONlt0q2lJnBwHnBc?m=slm-93jZN4A3soTozlEVIM4u
OsU6 LacO sgRNA	https://benchling.com/s/seq-vIyHY56Msirwh7nw742r?m=slm-ZR73jW5PrQLFUOOZOn6P
AtCLV3g1	https://benchling.com/s/seq-QWrmfZc1Ec4jfRGqrffm?m=slm-yvBAe0D1ae4A2FLpiaIH
AtCLV3g2	https://benchling.com/s/seq-asNJ6KJvO89swjwsrLNi?m=slm-vk1jze9mtjoP1TzoXRiw
FT BlockB sgRNA	https://benchling.com/s/seq-5ziJjbPSXUO7pftspKNn?m=slm-yPNQBDddE3vTFPT7e5wA
FT BlockC sgRNA	https://benchling.com/s/seq-PjFkpaHYbWxWtlo9kHf7?m=slm-dhtgd1CTfa1Clvdpywxk
FT BlockE sgRNA	https://benchling.com/s/seq-00MeZfV66n40zjyVIC9L?m=slm-vOxkbC9Faka3ax0Zt1Ts
FTgA sgRNA	https://benchling.com/s/seq-ZniKsGa2bnk8ZBummIJX?m=slm-bwtcu2gQfWNGLkHDabep
FTgB sgRNA	https://benchling.com/s/seq-qPKXU5P5KsZ4M9ekzaPX?m=slm-sjmjHtZt7ZGI5AqzW7qw
FT IntB sgRNA	https://benchling.com/s/seq-Lz9NPOLRl0r5fhwCDMUw?m=slm-CWTHhnhIBfr4TKMhU7hJ
FT IntC sgRNA	https://benchling.com/s/seq-5CjumRVX0cewolyPVvhQ?m=slm-esyIgOR1PRJkSNgaqKtk
FT IntE sgRNA	https://benchling.com/s/seq-UkhXeIHENHQ9mqmUaQLL?m=slm-sTSgfk7BIj0FNqndoEpg
PAP1 sgRNA	https://benchling.com/s/seq-gDuRWc02rOPJXxh6z9n8?m=slm-LlXCm27EI7u6OFjqeXKH
WUS sgRNA	https://benchling.com/s/seq-wk3d0aJUfeAMj9Hhfv5J?m=slm-Z1fhErOAFrgFMZzfX0zR

DRAFT

Supplementary Note 6: Supplementary Table S3. coding sequences

Name	Link
AtUbi10-dCas9-24xGCN4	https://benchling.com/s/seq-V82BZGxKbmZPsEALuIjz?m=slm-1VUiIIysQL7H07sPVKiB
CmYLCV-dCas9-24xGCN4	https://benchling.com/s/seq-pi1LDtQbPCT8AkOZA9zI?m=slm-6goMxLqKdLM5Fueln4Sf
p6xLacOfLuc	https://benchling.com/s/seq-sdCa6QHGpDqKMOJmuoxQ?m=slm-k8CSiyZg2MnRl3vEoSdX
pABI5	https://benchling.com/s/seq-tJH9toq3q0h5kmb0QFsZ?m=slm-hcIBNlydbsbwtjHqMEAs
pAda2b	https://benchling.com/s/seq-L4JonwyGnfLpC9v0p7ec?m=slm-AMxCq3zOQfdyJKo8Vnyt
pAFT1	https://benchling.com/s/seq-LE9lqv9JuA1uHB1QsyBC?m=slm-euBk5W2gIPNXvWZtWGQv
pAL2-TrAP	https://benchling.com/s/seq-ToKRHFff0aMbT075SfCl?m=slm-Jhx3STbRueBmY8qDcLQK
pAP1	https://benchling.com/s/seq-GND9iOIG4BvBqFy6RSOd?m=slm-xFfl4xOmu58VBl2NG8DA
pAsf1a	https://benchling.com/s/seq-hmnBk1OyBu9hzoNLl9nm?m=slm-QtifPCmgB4qQWpyT0exz
pAsf1b	https://benchling.com/s/seq-MYpDPxFoa4gKIS8dNRyR?m=slm-OxPcHnRB8QWTltm7XIez
pAtARF8	https://benchling.com/s/seq-CUyCV0yHwdn39DnX9dXn?m=slm-mnRDYOeryvBQ7rkYb0uZ
pAtARR2	https://benchling.com/s/seq-cNWHH2ierYKqPGig2e0q?m=slm-wErAczXmKCdE9GLiz1H2
pATHB13	https://benchling.com/s/seq-AASedSoGrBNOWDYNV66e?m=slm-Sl3u4Dg2oj97cboM8695
pAtHSFA2	https://benchling.com/s/seq-OzY54QANgYCK9k4kpHk4?m=slm-jmnSGI9SNhebcZ6qD8eC
pAtHSFA6b	https://benchling.com/s/seq-YKjpgyzzdrTIea4UkFZ7?m=slm-nr84ggjYXlxIa0o4PMns
pATXR7	https://benchling.com/s/seq-e1PrHK1m4z0ioVEim1ZE?m=slm-DJqzrV2QQLkRDuGLGqlF
pAvrXa10	https://benchling.com/s/seq-rJPtr5GAijv4WAH2tR9b?m=slm-203WHqlY6rhsamLtiCq7
pBBX21	https://benchling.com/s/seq-C4EDu8ltbug9MlnfDtKM?m=slm-79dpnpKf5csnCSLuZzVO
pbHLH122	https://benchling.com/s/seq-buRLTIFkqu2EAPfoDVBh?m=slm-4OgZoyHMmprl7LdDTIJ4
pBRE1	https://benchling.com/s/seq-ykNFHHaV7m1E8ORBfHLn?m=slm-cXwHxkoVVqvoS7HBlJse
pCVBp12	https://benchling.com/s/seq-Hq82z6HXPJ172OOkI9Lc?m=slm-OJsKRAu4TgTdUmLyqeNd
pDOF1	https://benchling.com/s/seq-N4SpuY16Zs8GttGdwBls?m=slm-b54eDByR6kaiuIyX4rO4
pDREB1	https://benchling.com/s/seq-7fAmR3xhXziwnjsd25OY?m=slm-cvgl5MtnNy7uSwalys7h
pDREB2	https://benchling.com/s/seq-kf27C99GuU8wOc3EzfAp?m=slm-lXGDP2cRbp2HJckBt8Ew
pDREB2-ZmUbi1	https://benchling.com/s/seq-TD5fXK41gOUJXps1xiA3?m=slm-gZKIeCKHdewH61B8OjFH
pDualLuc-35S	https://benchling.com/s/seq-elzbv9rmzIOLX0Z7N2Nr?m=slm-yJQeaA6mRVQ4ZDHlAeE3
pDualLuc-Nos	https://benchling.com/s/seq-u5wil7EGkWkB8JNNwudh?m=slm-WhnJihUzchH5egyIUzfy
pEFS	https://benchling.com/s/seq-oPwaKkVBQoWcH2bPtGqW?m=slm-TX6UkTfjFxo49ssqf04Q
pEIN3	https://benchling.com/s/seq-swdaXVh0CRDo42dYz9my?m=slm-E9bN7KhHneY5Gk7GW53Z
pGT3a	https://benchling.com/s/seq-aAZ4BrhfCyKhdYw49oaZ?m=slm-738xO8DT1a0GrJPDKPzM
pHAG1	https://benchling.com/s/seq-B2EAERI1sTtqEg4amng6?m=slm-ICBODhq1AwINVeX6Urpv
pIPN2	https://benchling.com/s/seq-Q6mY20wLg3hbcOniyd52?m=slm-oWckt6efGF9RzfciVID5
pLBD2	https://benchling.com/s/seq-ev22Ubl92DVWTDOFsuqR?m=slm-g79wFGa3hb0TNYLsQ2Ps
pLMI1	https://benchling.com/s/seq-KcSZODmVbyAaXL44f7Wl?m=slm-J1jhCIEpEPl2xvvXcX2M
pMYC2	https://benchling.com/s/seq-xs50vrNANfcsle9ZBWFv?m=slm-P7KuN1OKf5mfduEICpiz
pOpaque2	https://benchling.com/s/seq-yp9GBuPLSVrfHnROFZVW?m=slm-0aVbVlI6Pv6JJDQnSL6c
pOsCBT	https://benchling.com/s/seq-AamuHHSCGlxaG8HJL2d5?m=slm-YzpITtVPWz1B1isDZIFe
pOsGRF1	https://benchling.com/s/seq-EWHSTwdEJw1RuX0jhHT4?m=slm-sG63DEWVZbNzdwKhpKtK
pPTI4	https://benchling.com/s/seq-7Kd4fwQVcn7ozVbrcmYU?m=slm-7a8NlOTMIUKsoXG9qEyL
pRen-35S	https://benchling.com/s/seq-n2Xdx9MT7PwAekvj2BaJ?m=slm-OFczZX5aaj45PbIGsR2a
pRF2a	https://benchling.com/s/seq-QtvnHqeaayU3efKeWXr9?m=slm-8dsILuwyfaOawm9VKKge
pSDG31	https://benchling.com/s/seq-Q9bcGFZScL6VhTuTdm6M?m=slm-impam7UOTlt0zW2vmUKU
pSNAC1	https://benchling.com/s/seq-x8MispCo89a7Xt63olse?m=slm-x5VdYCjB5oZXKgoC1LMv
pTBP1	https://benchling.com/s/seq-xvVAE0mKixTQabsbmAVd?m=slm-z2y1tZDjNTyk1ERFBaea
pVRN1	https://benchling.com/s/seq-W320IrY6qyddeO89nld6?m=slm-o2cz2Pooiwl9IKy3uOGC
pWRKY50	https://benchling.com/s/seq-c7EdSUEJZL9Npl1g1NE8?m=slm-h4oxln4BGxYSfQveAbEc
pZmUbi1-dCas9-24xGCN4	https://benchling.com/s/seq-FjAPPyUNZMZsLAK7NJ8z?m=slm-UyTR5cy9mRRoIg9T0ALw
pZmVP1	https://benchling.com/s/seq-MsTk4ikHQdAEAgcwNaJO?m=slm-mClYrRWp5u7OVJMZYPWU

DRAFT

Supplementary Note 7: Supplementary Table S4. Binary vector maps

Name	Link
MT-AvrXa10-FT	https://benchling.com/s/seq-s8MMHPuv5O8ga4TWVYws?m=slm-zG7Srw29V6PMVhpTotUp
MT-DOF1-FT	https://benchling.com/s/seq-nVV28Wj4XlLoZvZzCZs1?m=slm-LttliC4AY8VRoTJ010CR
MT-DREB2-FT	https://benchling.com/s/seq-7rK40cVvEoHDMFPqs4LG?m=slm-UYlQe7OqoCpYZXcbzTJl

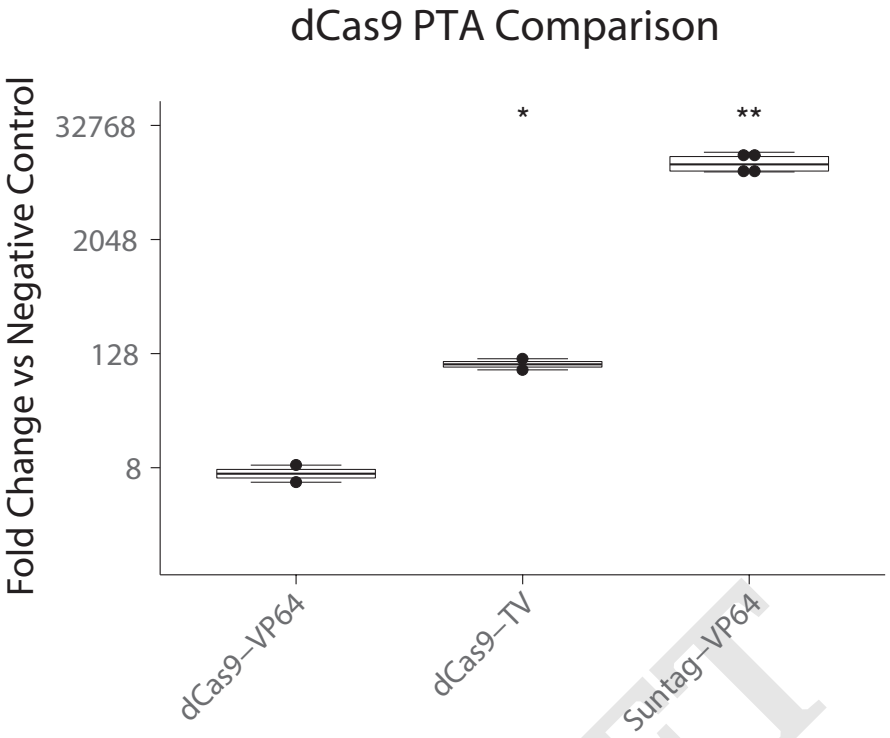
DRAFT

Supplementary Note 8: Supplementary Table S5. qRT-PCR primers

Name	Sequence (5'-3')
AtPP2A Fwd	AACGTGGCCAAAATGATGC
AtPP2A Rev	AACCGCTTGGTCGACTATCG
AtWUS Fwd	CAGCTTCAATAACGGGAATTT
AtWUS Rev	GTTGTAAGGTGCAGATGAGT
AtLEC1 Fwd	CAGAGAAACAATGGAACGTG
AtLEC1 Rev	TGAGACGGTAAGGTTTTACG
AtClv3 Fwd	GTTCAAGGACTTTCCAACCGCAAGATGAT
AtClv3 Rev	CCTTCTCTGCTTCTCCATTTGCTCCAACC
AtFT Fwd	CCCTGCTACAACTGGAACAAC
AtFT Rev	CACCCTGGTGCATACACTG
AtPAP1 Fwd	AGTATGGAGAAGGCAAATGGC
AtPAP1 Rev	CACCTATTCCCTAGAAGCCTATG

DRAFT

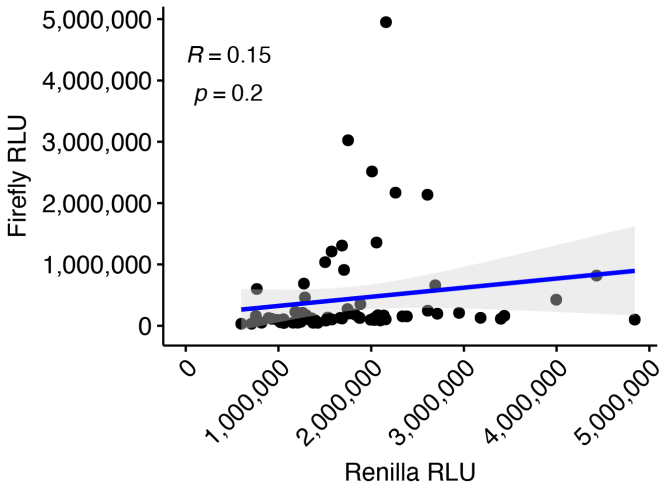
Supplementary Note 9: Supplementary Figure 1



Supplemental Figure 1. Direct fusion PTA vs SunTag PTA comparison in *A. thaliana* protoplasts. The direct fusion class of PTAs containing dCas9-VP64 and dCas9-TV, along with the SunTag-VP64 activator, were transformed into protoplasts with a 4x *Wuschel* sgRNA array targeting the core promoter for gene activation. RNA was extracted and activation was quantified by RT-qPCR using the delta delta Ct method. Parametric t-tests were performed comparing dCas9-TV or SunTag-VP64 with the dCas9-VP64 direct fusion at the same locus. * corresponds to p <= 0.05, ** corresponds to p <= 0.01, *** corresponds to p <=0.001, and **** corresponds to p<= 0.0001.

DRAFT

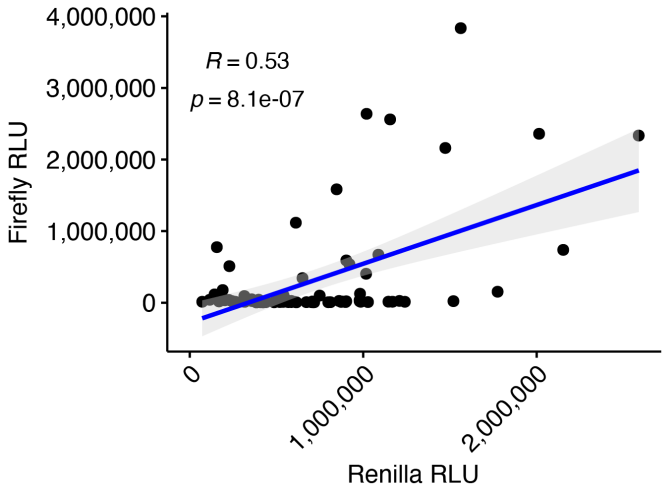
Supplementary Note 10: Supplementary Figure 2



Supplemental Figure 2. 35S Batch One *S. viridis*Dual Luciferase Assay The first protoplast isolation for *S. viridis* transformed with the library of scFv-ADs and the single dual luciferase vector. Renilla luciferase is driven by a 35S promoter. Data was fit to a linear regression model, with R and p values shown above.

DRAFT

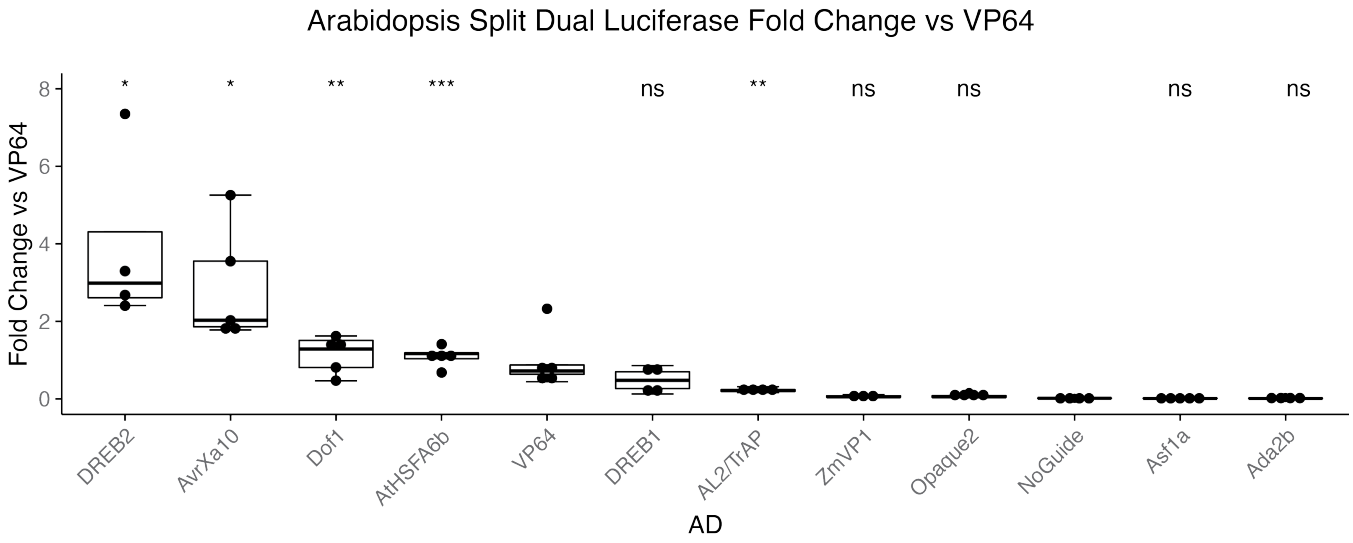
Supplementary Note 11: Supplementary Figure 3



Supplemental Figure 3. 35S Batch Two *S. viridis*Dual Luciferase Assay The second protoplast isolation for *S. viridis* transformed with the library of scFv-ADs and the single dual luciferase vector. Renilla luciferase is driven by a 35S promoter. Data was fit to a linear regression model, with R and p values shown above.

DRAFT

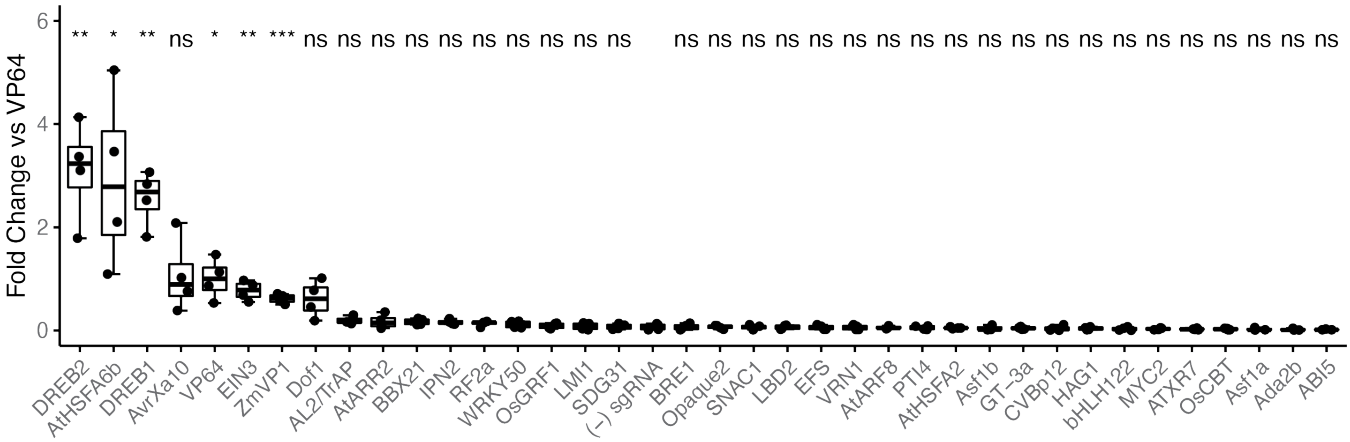
Supplementary Note 12: Supplementary Figure 4



Supplemental Figure 4. *Arabidopsis thaliana* Split Dual Luciferase Assay A smaller library of scFv-ADs was transformed into *A. thaliana* protoplasts along with the split dual luciferase vectors. The Firefly luciferase vector contained the 6xLacO-mini35S promoter, while the Renilla luciferase vector was driven by the constitutive 35S promoter. Firefly luciferase RLUs were divided by Renilla luciferase RLUs to normalize for transformation. The scFv-AD normalized RLUs were then divided by the scFv-VP64 normalized RLU to generate the plot show above.

DRAFT

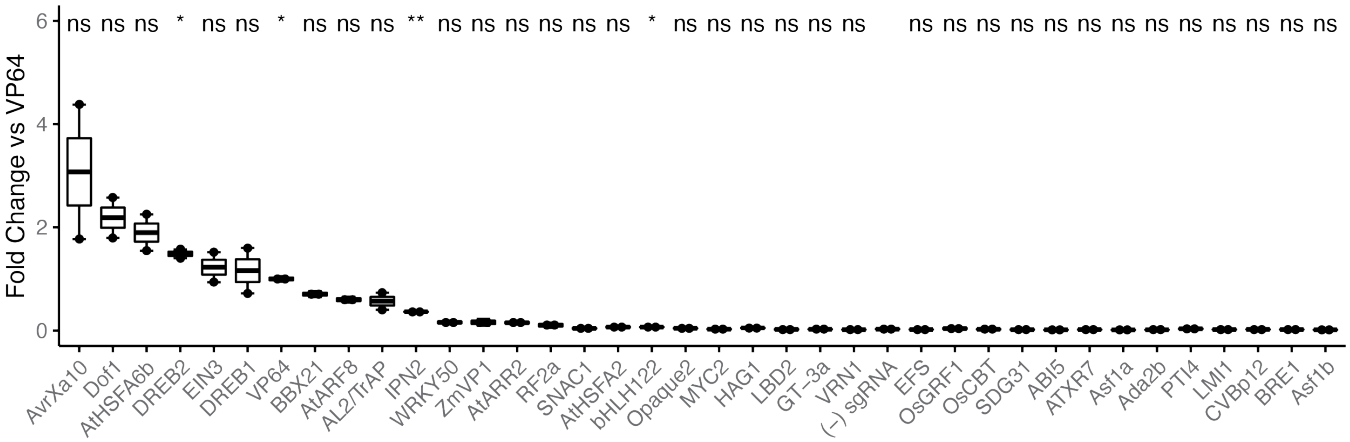
Supplementary Note 13: Supplementary Figure 5



Supplemental Figure 5. *Setaria viridis* Single Dual Luciferase Assay The full library of scFv-ADs was transformed into *S. viridis* protoplasts along with the single dual luciferase vectors. The single luciferase reporter contained the Firefly luciferase driven by the 6xLacO-mini35S promoter, while the Renilla luciferase was driven by the constitutive 35S promoter. Firefly luciferase RLUs were compared to the positive control VP64. A parametric t-test was performed comparing the negative control (-) sgRNA with each respective AD. p <= 0.05, ** corresponds to p <= 0.01, *** corresponds to p <=0.001, and **** corresponds to p<= 0.0001.

DRAFT

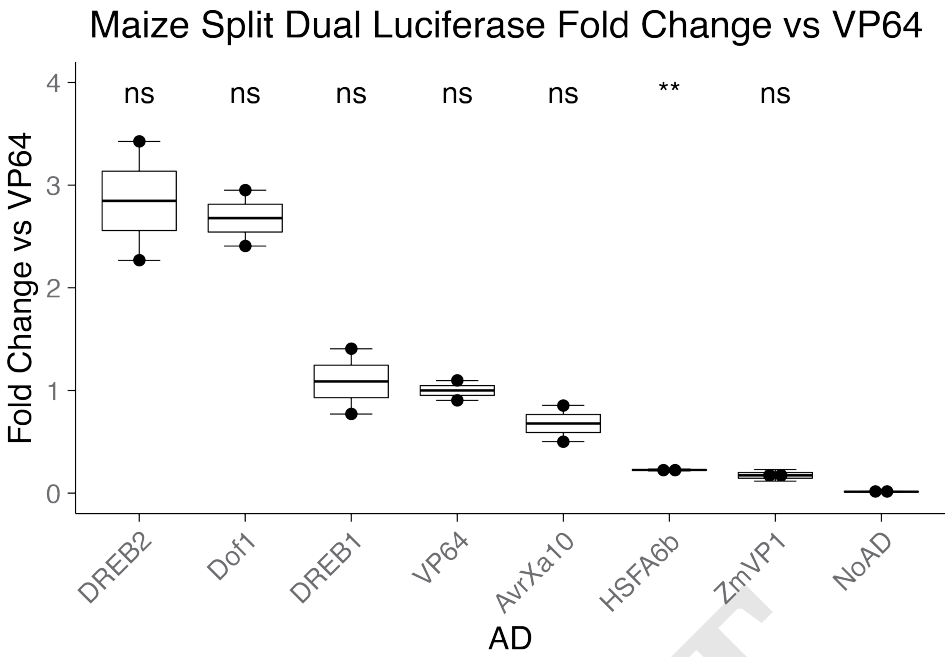
Supplementary Note 14: Supplementary Figure 6



Supplemental Figure 6. *Arabidopsis thaliana* Single Dual Luciferase Assay The full library of scFv-ADs was transformed into *A. thaliana* protoplasts along with the single dual luciferase vectors. The single luciferase reporter contained the Firefly luciferase driven by the 6xLacO-mini35S promoter, while the Renilla luciferase was driven by the constitutive 35S promoter. Firefly luciferase RLUs were compared to the positive control VP64. A parametric t-test was performed comparing the negative control (-) sgRNA with each respective AD. p <= 0.05, ** corresponds to p <= 0.01, *** corresponds to p <=0.001, and **** corresponds to p<= 0.0001.

DRAFT

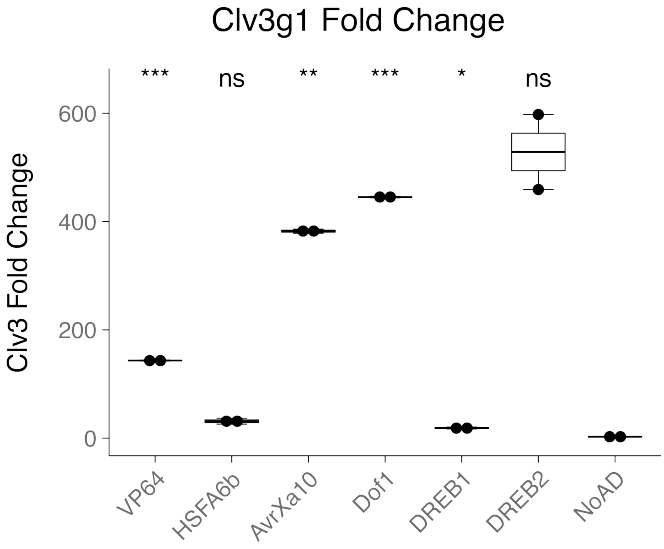
Supplementary Note 15: Supplementary Figure 7



Supplemental Figure 7. *Zea mays* Split Dual Luciferase Assay Greened *Zea mays* plants were used for protoplast isolation and transformation. A smaller library of scFv-ADs was transformed into *Z. mays* protoplasts along with the split dual luciferase vectors. The Firefly luciferase vector contained the 6xLacO-mini35S promoter, while the Renilla luciferase vector was driven by the constitutive 35S promoter. Firefly luciferase RLUs were divided by Renilla luciferase RLUs to normalize for transformation. The scFv-AD normalized RLUs were then divided by the scFv-VP64 normalized RLU to generate the plot show above.

DRAFT

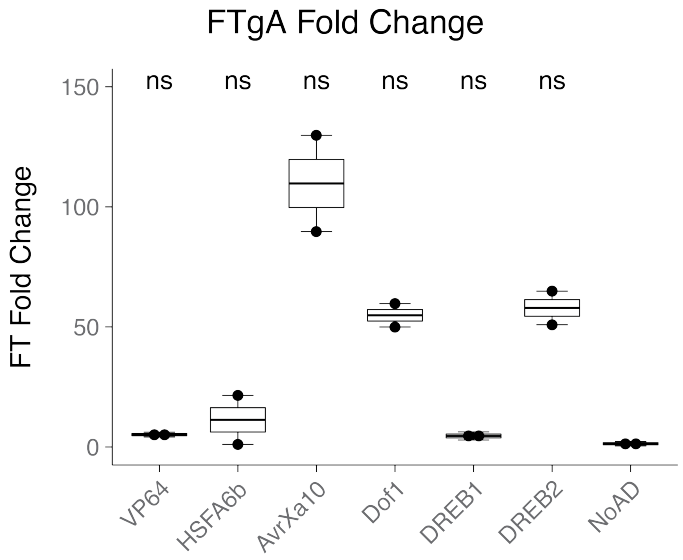
Supplementary Note 16: Supplementary Figure 8



Supplemental Figure 8. *A. thaliana* protoplast *Clv3* gene activation with sgRNA1 Protoplasts were isolated from *A. thaliana* and transformed with Ubi10-dCas9-24xGCN, 35S-scFv-AD, and AtU6-Clv3g1. RNA was isolated from cells 24h post-transformation, and RT-qPCR was performed to quantify gene expression. The delta delta Ct method was used to calculate Fold Change vs the NoAD negative control. PP2A was used as the housekeeping gene.

DRAFT

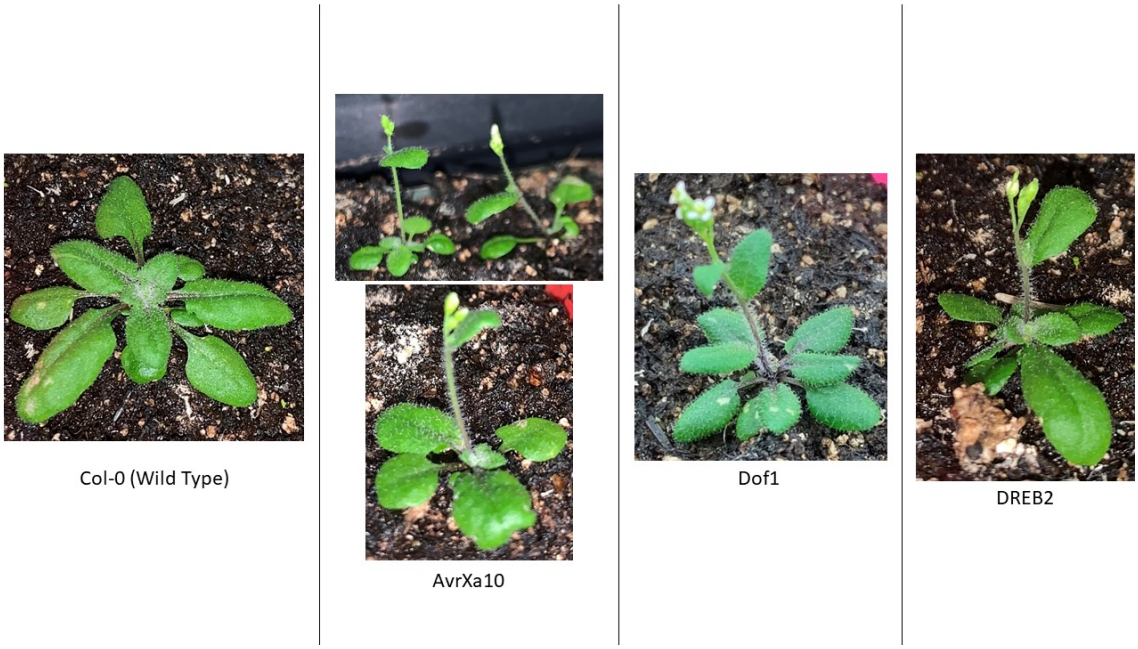
Supplementary Note 17: Supplementary Figure 9



Supplemental Figure 9. ***A. thaliana* protoplast *FT* gene activation with FTgA** Protoplasts were isolated from *A. thaliana* and transformed with Ubi10-dCas9-24xGCN, 35S-scFv-AD, and AtU6-FTgA. RNA was isolated from cells 24h post-transformation, and RT-qPCR was performed to quantify gene expression. The delta delta Ct method was used to calculate Fold Change vs the NoAD negative control. PP2A was used as the housekeeping gene.

DRAFT

Supplementary Note 18: Supplementary Figure 10



Supplemental Figure 10. T1 *Arabidopsis thaliana* Parental Lines Images showing the T1 parental lines and an age-matched wild type Col-0 *A. thaliana* plant. The binary vectors for these plants are identical, except for the AD fused to NbGP41. Labels below each image denote the AD in each parental line.

DRAFT