ENGINEERING GENE EXPRESSION IN PLANTS WITH PROGRAMMABLE TRANSCRIPTIONAL ACTIVATORS

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

MATTHEW HARLEY ZINSELMEIER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DANIEL F VOYTAS, PHD

ADVISOR

MICHAEL J SMANSKI, PHD

CO-ADVISOR

MAY, 2022

{Copyright page}

Acknowledgements

Thank you to my advisors Dan and Mike. Dan – you foster an amazing lab space for us to grow as students, scientists, and professionals. Your big-picture outlook is unique and conversations with you about my research always brought a fresh perspective that motivated me to continue pushing onward. Mike – While Dan provides the big picture, I learned about the details from you. Starting with our bi-weekly meetings, I learned the true nitty gritty details of molecular biology. Your mentorship has molded my work ethic, planning, and approach towards the day-to-day work of being a scientist. I look up to both of you greatly and that will never change. I am forever grateful to you both for taking the chance on a kid from Iowa. Thank you.

I would like to acknowledge the wonderful friends I made at the University of Minnesota over the course of my graduate school career. Notable individuals with whom I shared scientific discussions within the Voytas and Smanski Labs were Joe Belanto, Tomas Cermak, Eva Konecna, Qiwei Shan, Nat Graham, Arjun Khakhar, Jon Cody, Paul Atkins, Michael Maher, James Chamness, Redeat Tibebu, Maria Elena Gamo, Maciej Maselko, Stephen Heinsch, and Dimitri Perusse. I would like to specifically recognize Juan 'Armando' Casas-Mollano, Ryan Nasti and Evan Ellison for going above and beyond in their friendship over the years, whether it was results in the lab or just day-to-day survival as a graduate student. In addition to colleagues in the lab, I would like to acknowledge Eric Aird, Em Lopresti, Colette Rogers, and Alina Zdechlik as friends from my cohort who I leaned on for social interaction outside the lab.

Most importantly is my family. My grandparents Beth and Erv lived in Roseville, MN for the duration of my PhD career and always helped when I needed family in the area. My grandparents Bob and Ellie Zinselmeier were never bashful about cheering me on from afar. Bob and Erv, I miss you both dearly. May you rest in peace. My brother and sister Tom and Anna Zinselmeier are the best a brother could ask for; whether it

was spontaneous trips with my sister or just hanging out with my brother on the weekend, you were both there for me when I needed you. My parents Chris and Sharon Zinselmeier. I could write a thesis chapter about how much you mean to me, and I could not have done any of this without you. Last and certainly not least is Anna Libra. We met my first year in the program. You have been at my side ever since. I love you.

To my father Chris and my late grandfather Erv, you are my inspiration. You are and were two incredible scientists I am always striving to match. I wouldn't be where I am today without you both.

Abstract

Agriculture is a centuries old practice that has selected upon natural variation over time. Highly productive cultivars today are the result of this selection. DNA sequencing has revealed the genetic blueprint for many of these crop species, allowing for precise selection of variants for breeding. Crops must survive and reproduce efficiently to be utilized by humans for agriculture. To accomplish this, crops put the DNA blueprint into action through gene expression, allowing for development and survival in the face of stress. Thus, understanding and controlling gene expression will be important for engineering highly productive crops around the world.

In nature, transcription factors (TFs) are responsible for regulating gene expression. TFs are comprised of a DNA binding domain and transcription regulatory domain. The DNA binding domain will bind to a region in the genome, while the regulatory domains interact with other proteins capable of either initiating or blocking transcription. As programmable nuclease technology like CRISPR-Cas9 was elucidated, Programmable Transcription Activators (PTAs) were developed to function as engineered transcription factors controlling gene expression. PTAs can be targeted anywhere in the genome to activate expression of a target gene promoter. PTAs can also activate the expression of multiple genes at once. In Chapter One the status of PTA technology is reviewed, with systems showing promise in plant backgrounds given consideration. To ultimately use PTAs for basic plant research, we first set out to optimize PTAs for efficient performance across plant species.

The VP64 activation domain is the most frequently used activation domains regardless of PTA system. This domain is derived from a human herpes virus yet is still used in many plant systems. To address this we designed, built, and tested a library of plant-derived activation domains for activity in plant cells, as described in Chapter Two. The AvrXa10, Dof1, and DREB2 ADs proved to be efficient across a variety of plant

species. We also demonstrate the use of the DREB2 AD to activate distal enhancers in *A. thaliana* protoplasts, showcasing the versatility of plant-optimized PTAs.

Finally, PTAs were used to engineer a circuit for genetic biocontainment as described in Chapter 3. In certain crop species, there is a desire to prevent gene flow between closely related individuals. A system was devised called Engineered Genetic Incompatibility (EGI) to prevent crop transgenes from escaping into closely related weedy populations. To engineer this system a gene is identified that is capable of producing lethality when overexpressed using PTAs. Gene editing is then performed to remove the PTA binding site to comprise the EGI organism; a PTA is expressed alongside the modified promoter within an individual to prevent self-targeting. Upon crossing with a wild-type individual, hybrid progeny will contain one copy of the PTA and one copy of the unedited promoter resulting in lethal overexpression. Our efforts to implement EGI in the model organism A. thaliana are described in detail including selection of target genes, testing sgRNA activity, generating transgenic lines, and promoter mutagenesis. The work described in this thesis illustrates the development and application of PTA technology in agriculture to engineer more productive crops through controlling gene expression.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

Chapter 1

CRISPR-Cas Activators for Engineering Gene Expression in Higher Eukaryotes

J. Armando Casas-Mollano,1 Matthew H. Zinselmeier,2 Samuel E. Erickson,1 and Michael J. Smanski1,*

1 Department of Biochemistry, Molecular Biology, and Biophysics, BioTechnology Institute, University of Minnesota, Twin-Cities, Saint Paul, Minnesota, USA; and 2 Department of Genetics, Cell Biology, and Development, University of Minnesota, Twin-Cities, Saint Paul, Minnesota, USA. *Address correspondence to: Michael J. Smanski, PhD, Department of Biochemistry, Molecular Biology, and Biophysics, BioTechnology Institute, University of Minnesota, Twin-Cities, 1479 Gortner Ave., Room 140, Saint Paul, MN 55108, USA, E-mail: smanski@umn.edu

Re-printed with permissions

1.1 Abstract

CRISPR-Cas-based transcriptional activators allow genetic engineers to specifically induce expression of one or many target genes in trans. Here we review the many design variations of these versatile tools and compare their effectiveness in different eukaryotic systems. Lastly, we highlight several applications of programmable transcriptional activation to interrogate and engineer complex biological processes.

1.2 Introduction

The engineering of CRISPR-Cas (CRISPR-associated) systems has provided a means for simple, accurate, and efficient genome editing. The type II CRISPR-Cas9 system from *Streptococcus pyogenes* is the most commonly used for genome editing.[1] This ribonucleoprotein complex consists of a DNA endonuclease (Cas9) and two RNAs, CRISPR RNA (crRNA) and transacting RNA (tracrRNA). Together, these components make a blunt cut in DNA upon binding to the target sequence.[2–4]

The crRNA component of the complex provides sequence specificity by base-pairing to the complementary 20 nucleotides of the target DNA (protospacer) upstream of an ‘‘NGG’’ protospacer adjacent motif (PAM).[3,4] crRNA and tracrRNA can be combined into a small guide RNA (sgRNA) that is sufficient for function[4]. Thus, the two-component sgRNA-Cas9 complex constitutes an RNA-guided platform for cleaving specific genomic regions. This tool has transformed the cost and throughput of genome editing in recent years.

CRISPR-Cas9 systems have been engineered to remove their DNA cutting activity for applications that exploit their sequence-programmable DNA-binding ability. Cas9 contains two endonuclease domains, HNH and RuvC, each involved in the cleavage of a single DNA strand.[4] Thus, to create a catalytically inactive or ‘‘dead’’ Cas9 (dCas9), both domains were inactivated[4]. Later, it was demonstrated that dCas9 proteins

harboring mutations that inactivate the endonuclease catalytic sites retain their DNA binding abilities.[5,6] Qi and colleagues used a dCas9 with D10A and H840A substitutions in the RuvC and HNH domains, respectively[4,5]. A second dCas9 version was designed using four mutations in the nuclease domains, D839A and N863A in addition to the abovementioned D10A and H840A.6 Both dCas9 versions are devoid of nuclease activity but remain strong RNA-guided DNA binding proteins.[5,6]

dCas9 has been fused to many effector domains with the goal of recruiting different activities to locally modify the target DNA or its associated proteins. When transcription activation domains (ADs) are fused to dCas9, the resulting protein can induce expression of genes in the vicinity, thus becoming a programmable transcriptional activator (PTA), also known as a CRISPR activator (CRISPRa). This review focuses on PTAs created using CRISPR-Cas systems, describing the different designs, common features, and *in vivo* applications. We focus on PTAs developed and tested in eukaryotic organisms.

1.3 Activation Domains

ADs are defined as motifs capable of recruiting the transcription preinitiation complex (PIC) to a core promoter. RNA polymerase II and the general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH comprise the PIC.[7,8] The mediator, a large complex of reversibly associating transcriptional regulatory subunits, is part of the PIC through interactions with RNA polymerase II.[9] Strong ADs interact with components of the PIC, accelerating its assembly at a core promoter.[10] The mechanism by which these interactions occur relates to the conserved architecture of many ADs.

ADs commonly used in conjunction with dCas9 programmable DNA binding domains are shown in Table 1. VP64, a common AD, is a tetrameric repeat derived from the VP16 protein of herpes simplex virus.[11] Fusing four end-to-end repeats of the VP16 motif (VP64) enhanced its ability to activate transcription.[12] Furthermore, fusion of the

VP64 domain to additional ADs resulted in even greater transcription activation.[13,14] The ADs from p65 (nuclear factor kappa B, 65 kDa subunit) and Rta (Epstein–Barr virus R transactivator) are two of the domains commonly used as a fusion to VP64.

Transcriptional activation mediated by p65 is conferred by two distinct C-terminal transactivation domains, TA1 and TA2.[15] Similarly, Rta contains two acidic C-terminal activation subdomains, 1 and 2, from which domain 2 confers the most potent transcriptional activity.[16] Another domain used in the engineering of PTAs in animals is from human heat shock factor 1 (HSF1). HSF1 AD comprises regions B and C located at the C-terminus of the protein.[17]

ADs such as EDLL (APETALA2 family activation domain) and CBF1 (C-REPEAT/DRE BINDING FACTOR 1) are derived from plant species and activate target genes in both plant and mammalian cells.[18–20] EDLL is characterized by a distinctive distribution of acidic residues and hydrophobic leucines located at the C-terminus of AP2/ERF family of plant transcription factors.19[19] The AD of CBF1 comprises acidic amino acid residues located at the C-terminal half of the protein.[18]

Transcription activator-like acidic activation domain (TAL) is a prokaryotic AD from the C-terminal region of the transcription activator-like effector (TALE) protein, a *Xanthomonas* transcription factor secreted into plant cells to regulate gene expression in the host.[21] The TAL AD was demonstrated to been able to induce expression of target genes in yeast and plant cells.[21]

ADs commonly used in PTA systems (Table 1) tend to be intrinsically disordered motifs. ADs have previously been classified as ‘‘acidic blobs’’ or ‘‘negative noodles’’ enriched in acidic, proline, serine, threonine, and glutamine residues.[22] This trend is described by a model in which disordered acidic ADs recruit coactivators by concurrent AD-coactivator phase separation at a promoter.[10] This mechanism relies upon the propensity for intrinsically disordered regions to form scaffold-like structures by exposing

short repeating peptide motifs along with conformational flexibility.[23] ADs, coactivators, and RNA Pol II itself contain intrinsically disordered low complexity regions.[24]

The following model has emerged for dynamic and reversible assembly of the PIC. Genomic targeting of an intrinsically disordered AD, which can interact with coactivators, forms a phase-separated condensate at a core promoter. This promotes clustering of other intrinsically disordered proteins at the droplet such as additional coactivators, general transcription factors, and RNA Pol II CTD, resulting in an active PIC.[25]

1.4 CRISPR-Cas Activation Systems

1.4.1 - First-generation PTAs

The first CRISPR-based transcriptional activators were created by fusing an AD to the C-terminus of dCas9 (Fig. 1 and Table 1). The AD of choice was VP16 in any of its multiple iterations (four tandem repeats, VP64 or 10 repeats, VP160) (Fig. 1a). The chimeric dCas9-VP64 is able to activate transcription of reporter and endogenous genes when targeted to promoter regions.6,26–31[6,26–31] Transcriptional activation of some genes led to the increased accumulation of endogenous protein.[29–31] However, most targeted genes showed only modest to low levels of transcriptional activation.[6,28,31] Even for dCas9 constructs with the stronger activator VP160, only 10-fold activation was observed.[26]

Other direct fusion PTAs have been created. For instance, a dCas9-p65 fusion is capable of transcription activation although to lower levels than dCas9-VP64.28 In plants, dCas9 fused to EDLL, TAL, and CBF1 activator domains, all increase gene activation at significant, but still modest levels.[20,32]

1.4.2 Second-generation PTAs

The second generation of CRISPR-based activators leverage the synergistic effect that the recruitment of multiple transcription factors has in natural activating

systems.[13] Thus, the main premise is that recruiting multiple ADs to the promoter, as single repeated domains, heterogeneous combinations, or both, will enhance transcriptional activation. This has been achieved through diverse and creative mechanisms (Table 1).

1.4.3 - dCas9-VPR

In the dCas9-VPR system, the efficiency of transcriptional activation was greatly improved by making tandem fusions of different ADs (Fig. 1b). First, to identify suitable ADs, Chavez et al. screened 22 single AD dCas9 C-terminal fusions for their ability to activate transcription of a reporter.[13] VP64, p65, and Rta were the strongest ADs.[13] To increase the strength of the transcriptional activity, tripartite activators were created by sequential fusion of the ADs to the C-terminus of dCas9 in different orders.

The fusion providing the highest transcriptional activation was VP64-p65-Rta, VPR for short. dCas9-VPR performs significantly better than dCas9-VP64, any of the single ADs fused to dCas9, and constructs containing double AD fusions.[13] dCas9-VPR can activate transcription of endogenous genes in animal cell lines, in some cases to levels comparable with those observed in native tissues. Furthermore, activation of *neurogenin 2* or *neurogenic differentiation factor 1* by sgRNA-guided dCas9-VPR was robust enough to induce differentiation of human-induced pluripotent stem cells into induced neurons, which was not possible using dCas9-VP64.13[13] Thus, dCas9-VPR constitutes a strong transcriptional activator capable of inducing gene expression to levels meaningful enough to exert phenotypic changes.

1.4.4 - dCas9-TV

dCas9-TV was similarly developed by fusing tandem repeats of ADs to the C-terminus of dCas9 (Fig. 1c). Increasing the number of AD repeats led to an increase in transcriptional activation. However, too many repeats fused to dCas9 led to decreased protein accumulation, possibly due to instability triggered by the repetitive nature of the

constructs.[14] By testing different combinations of four ADs, Li et al. created an efficient PTA with a number of AD repeats that balances PTA stability and target gene overexpression.[14] This optimized activator, named dCas9-TV, is a fusion of dCas9 to six tandem copies of TAL followed by eight copies of VP16 (Fig. 1c).[14]

dCas9-TV was tested in cells from eudicot and monocot plants and significantly activates transcription of endogenous genes. dCas9-TV is also capable of activating *ASCL1* and *OCT4* in HEK 293T human cells. Transgenic plants expressing dCas9-TV and an sgRNA targeting the promoter *RLP23*, a leucine-rich repeat receptor protein that mediates immune response, displayed enhanced immune response in the presence of the peptide elicitor, nlp20, of RLP23.[14]

1.4.5 - Scaffold RNA and synergistic activation mediator

Scaffold RNA (scRNA) and synergistic activation mediator (SAM) are based on engineering hairpins in the sgRNA structure that allow the use of RNA-binding proteins to tether ADs to the dCas9 ribonucleoprotein complex (Fig. 1d, e). In the scRNA system, two MS2 RNA hairpin loops are covalently attached to the 3' end of the sgRNA. MS2 hairpin loops are bound specifically and avidly by dimers of the MS2 bacteriophage coat protein (MCP). When a chimeric MCP-VP64 protein is coexpressed with the modified sgRNA and dCas9, the assembled ribonucleoprotein complex will contain up to four copies of the VP64 domain (Fig. 1d).[6]

A similar approach was used in the SAM system, however, in this case, the MS2 RNA hairpin loops are appended to the tetraloop and stem-loop 2 of the sgRNA to create the so-called sgRNA2.0.[33] Since both the tetraloop and stem-loop 2 extend beyond the surface of the dCas9- sgRNA complex, the addition of the hairpins does not affect the DNA-binding of dCas9 and still allows recruitment of MCP-AD fusions.[33] In addition, to increase the effectiveness of the SAM system, p65 and HSF1 ADs were fused in tandem to MCP and a single VP64 AD was directly linked to dCas9 (Fig. 1e).[33]

SAM is a potent transcriptional activator, consistently outperforming dCas9-VP64.[33] Conversely, the initial iteration of scRNA was up to three times less effective than dCas9-VP64.6 The low performance of the first iteration of the scRNA system may be due to the instability of the modified sgRNA caused by the addition of multiple repeats of the MS2 loop.[34] Indeed, redesign of the multi-hairpin to improve stability resulted in a more potent activator.[34]

scRNA and SAM can be engineered as platforms for broad control of gene expression. The use of orthogonal sets of RNA-hairpins:binding-proteins produces distinct regulons. sgRNAs that recruit different effector domains to the dCas9 protein, will confer unique effects at each sgRNA target locus.[6,34] This flexibility allows for the creation of complex sgRNA-encoded programs using dCas9 as a master regulator.[34]

1.4.6 - SunTag

In the SunTag system, multiple copies of an AD are targeted to the dCas9 ribonucleoprotein through an epitope/antibody interaction.[35] The SunTag PTA is composed of an sgRNA and two protein modules: (1) dCas9 protein fused to a tandem array of GCN4 epitope motifs separated by flexible GS linkers, and (2) a singlechain variable fragment antibody (ScFv), with affinity for the GCN4 motif, fused to VP64. Upon expression of both protein components, the GCN4 epitope array on dCas9 recruits up to 10 copies of the ScFv-VP64 AD. This complex is targeted to the promoter-of-interest by the sgRNA (Fig. 1f).[35]

SunTag is a powerful activator of gene expression. When introduced into mammalian cells, the SunTag system increased expression of target genes and produced the expected phenotypes.[35] In contrast to the results obtained with SunTag, activation of these same genes by dCas9-VP64 was very inefficient and did not yield the expected cell responses.[35] Thus, transcriptional activation by SunTag is robust enough to produce the biological response predicted by the overexpression of a target gene.

1.4.7 - Three-component repurposed technology for enhanced expression

The three-component repurposed technology for enhanced expression (TREE) system enhances the recruitment of multiple AD copies via a hierarchical multitag system. It combines SunTag with the RNA tethering system used by SAM in a tree-resembling architecture (Fig. 1g).[36] The primary tag is the RNA hairpin loop bound by an MCP-GCN4-array fusion protein. Finally, the AD is recruited to the complex by fusion to an ScFv with affinity for GCN4. Both the p65-HSF1 and tripartite VPR ADs have been used in the TREE system (Fig. 1g).[36]

Like other second-generation PTAs, the TREE system gives strong transcriptional activation. TREE outperformed dCas9-VP4, the SAM system, and dCas9-VPR when using p65-HSF1 and VPR as ADs, respectively. A direct comparison with the canonical 10-copy SunTagVP64 was not published, but TREE appears to outperform a SunTag version recruiting up to four and eight copies of VP64.[36]

1.4.8 - Engineering PTAs using novel CRISPR-Cas systems

Most PTA designs currently available for RNA-guided activation use the *S. pyogenes* dCas9 ribonucleoprotein complex. However, characterization of alternative CRISPR systems has provided researchers with a broader set of CRISPR-Cas proteins for engineering of novel PTAs (Table 1).

One of these proteins is Cpf1 (CRISPR from *Prevotella* and *Francisella 1*), also known as Cas12a, an RNA-guided endonuclease producing staggered DNA double-stranded breaks.[37] A DNase-dead Cpf1 (ddCpf1) that retains the sgRNA-guided DNA binding function was generated by inactivating the RuvC domain.[38] ddCpf1 fused to VP64 or p65 ADs activates transcription of target genes, although not very efficiently.[39,40] However, more robust gene activation was obtained when ddCpf1 was used in ddCpf1-VPR and ddCpf1-SunTag systems.[39–41] Conversely, replacing dCas9 by ddCpf1 in the TV system yielded only a weak activator.[14] Attempts to create the ddCpf1 equivalent of

SAM or scRNA were unsuccessful because the stem loop region and the 3' end of the crRNA do not protrude from the Cpf1 ribonucleoprotein complex.[42]

Recently, a design not first demonstrated with dCas9 was used to create a novel ddCpf1 activator. Nihongaki et al. generated a split form of ddCpf1 that spontaneously associates to yield a functional heterodimer.[42] This split form duplicated the number of N- and C-terminal ends available for attaching ADs.[42] The bipartite AD, p65- HSF1, was fused to each end of the two halves of ddCpf1 (Fig. 1h). This split ddCpf1 PTA was able to activate multiple endogenous genes in HEK293T cells. Compared with the dCas9-based SAM system, split ddCpf1 consistently reached higher activation levels.[42]

In addition to the effectiveness of ddCpf1 for the design of potent PTAs, this protein possesses additional features that make its use compelling: (1) binding specificity for Cpf1 to its DNA target is greater than for Cas9[43–45]; (2) the Cpf1 ribonucleoprotein complex comprised a crRNA while lacking the tracrRNA, simplifying the design of the sgRNAs[37]; (3) the T-rich PAM used by Cpf1 enables the targeting of promoter regions not covered by the G-rich PAM of Cas9[37]; and (4) the endogenous RNase activity of Cpf1 simplifies the generation of multiple crRNAs from the processing of a single transcript carrying a crRNA array.[46]

Both Cas9 and Cpf1 are Class 2 CRISPR systems that utilize a single-protein component for nuclease activity. This is in contrast to class 1 systems that feature multicomponent nucleases.[47] Despite the added complexity, the diversity of class 1 CRISPR-Cas systems offers some advantages for the development of PTAs. Young et al. have reported a novel PTA based on a Class 1/ type I-E complex from *Streptococcus thermophilus* DGCC7710 called Cascade (CRISPR-associated complex for antiviral defense).[20] Type I-E Cascade comprised 6 subunits (Cas3, CasA, CasB, CasC, CasD, and CasE) from which CasE is involved in crRNA processing, CasABCD in target recognition, and Cas3 as a single-stranded exonuclease.[20]

Since Cas3 is recruited only after the complex is poised at its target sequence, a DNA binding complex could be obtained by simply excluding Cas3 from the system. To fashion a Cascade PTA, the CBF1 AD was fused to the C-terminal end of CasA, CasD, and CasE (Fig. 1i). The resulting PTA, Sth Cascade, activated transcription of a reporter and an endogenous target gene (r) when transiently expressed in maize embryos. Activation of the transcription factor R by Sth Cascade was robust enough to produce the expected anthocyanin phenotype. However, it was not significantly better than that obtained using dCas9-CBF1.[20]

1.5 Common Features of Gene Activation Mediated by CRISPR Activation Systems

Despite the diversity of PTA designs, target genes, and experimental organisms, some features common to all systems have started to emerge (Fig. 2).

1.5.1 - Synergy

Before the discovery of CRISPR-Cas systems, it was established that many promoters contain multiple binding motifs for the same transcription factor. When multiple copies of a transcription factor are recruited to a promoter, they interact synergistically to enhance transcriptional activation.[48–50] Thus, early in the inception of CRISPR-based PTAs, targeting multiple regions of a gene was explored as an approach to increase potency. Certainly, when multiple dCas9-VP64 complexes are recruited to a promoter, they act in concert to yield a stronger transcriptional activation (Fig. 2a).[6,30,31] This same observation holds true for most second-generation PTAs and across multiple organisms.

This effect had been previously reported with TALE-based PTAs.[51] However, an advantage of CRISPR-Cas-based PTAs is that DNA targeting is mediated by complementarity to a crRNA (or sgRNA). This facilitates targeting multiple sites upstream of a gene-of-interest. Doing the same with TALEs would require engineering and expressing multiple 4 kb transgenes in the same construct.

1.5.2 - Position effects

sgRNA targeting of a PTA to different regions of a gene, in any orientation, may confer some degree of gene activation.[30,52] However, the strength of overexpression is influenced by the proximity of PTA binding to the transcription start site (TSS). Systematic screens determined that binding in the upstream region close to the TSS induces the strongest gene activation, although to unpredictable levels (Fig. 2b).[6,26]

Targeting the dCas9-VP160 activator using sgRNAs binding the region 300 bp upstream of the TSS of *IL1RN*, *SOX2*, and *OCT4* was most effective, whereas using sgRNAs binding the region downstream of the TSS had adverse effects on activation.[26] Similarly, the most potent sgRNAs for dCas9-VP64 targeted a window 147–89 bp upstream of the TSS of the mouse OCT4 gene.[53] For the SunTag activator, sgRNAs showing the highest activation bound 400–50 bp upstream of the TSS.[52] For the targeting of the SAM activator, the strongest induction of expression was obtained with sgRNAs located within 200 bp upstream of the TSS of 12 human genes.[33]

In yeast cells, robust activation was obtained when dCas9-VPR was targeted within 400 bp upstream of the start codon, yet the efficiency decreased when the sgRNAs were located within 20 bp of the TATA box.[54] In rice plants, binding of dCas9-TV to the region within 300 bp of the TSS of *OsWOX11* and *OsYUC1* yielded the strongest gene activation responses.[55] Similar targeting windows were found for TALE-based PTAs.[53] Consequently, hitting this ‘‘sweet spot,’’ a couple of hundred base pairs upstream of the TSS, should be a general consideration when designing optimal sgRNAs. It is unknown how optimal targeting sites will covary with PTA designs, an important point when considering studies that compare diverse PTA designs at a single binding site.

1.5.3 - Gene to gene variability in relative efficiency

dCas9-VP64-mediated activation of gene expression in the human cell line HEK293T was observed using one sgRNA for some genes, while others required the synergistic activity of pooled sgRNAs.[31] A similar gene-to-gene variability in the relative strength of transcriptional activation was observed with dCas9-VPR and SAM systems.[13,33] The chromatin state around the promoter was thought to cause this variability, but open chromatin, as determined by the presence of DNase I hypersensitivity sites, is not a prerequisite for gene activation.[31]

A negative correlation exists between basal gene expression and relative fold-change in activation conferred by CRISPR-Cas activators (Fig. 2c).[13,33,56] In animal cells, this correlation was observed in all dCas9-based activators tested, dCas9-VP64, dCas9-VPR, SAM, and SunTag.[56] Likewise, in Drosophila, the ability of dCas9-VPR to activate transcription of a given gene depended on its basal expression level.[57] In plants, a similar correlation exists for genes activated by dCas9-TV.[14]

Even though the relative change in gene expression induced by most PTAs will be larger for weakly induced genes, the absolute transcription rate that a gene may reach depends on the system used.[56] Thus, when activating a gene, its steady-state rate should be taken into account and the use of multiple sgRNAs considered when designing PTAs to achieve maximum gene expression.

1.5.4 - Multiplexing

One unique advantage of CRISPR-based PTAs is the ease of reprogramming the target site. This property allows for multiplexed gene activation by directing a single PTA to different targets via multiple coexpressed sgRNAs (Fig. 2d). Multiplexed regulation was first reported using dCas9-VP160. Transfection of dCas9-VP160 and sgRNAs targeting *SOX2*, *IL1RN*, and *OCT4* led to the concurrent activation of each gene.[26] Furthermore, the relative activation of the three genes could be modulated by changing the dosage of the individual sgRNAs.[26]

Multiplexed gene activation of up to 10 genes, using 10 sgRNAs, was achieved using the SAM system in mammalian cells.[33] However, this led to a global reduction in the absolute activation levels of each gene.[33] In addition, the relative activation efficiencies among genes change in single- versus multiplexed assays.[33] A similar reduction in overall activation was observed in the multiplexed activation of three genes, *twist*, *snail*, and *engrailed* using dCas9-VPR in Drosophila cells.[57]

These studies suggest that the concentration of dCas9 sets the upper limit for transactivation, and this resource is allocated between the various sgRNAs expressed in the cell. In contrast, another study did not find differences in efficiency between single and multiplexed activation of six target genes with dCas9-VP64, dCas9-VPR, SunTag, or the SAM system.[56] However, variability in the levels of basal expression of target genes makes direct comparisons of these experiments challenging.[56] Multiplexing CRISPR-based PTAs can be exploited for the manipulation of metabolic pathways and for the rewiring of gene expression networks to yield complex phenotypes.[34]

1.5.5 - Specificity

The specificity of CRISPR-Cas PTAs is a major concern for their use in living systems. Unintended gene expression changes caused by off-target binding may cause undesired effects and may even lead to reduced fitness or survival. Off-target effects have been documented with the Cas9 nuclease, which induced cleavage of up to five off-target sites in the human genome.[58,59]

In contrast, transcriptome-wide analysis by RNAseq demonstrated that CRISPR-Cas-based PTAs are very specific with little off-target effects. In human cells, gene activation mediated by the relatively weak activators dCas9-VP64 and dCas9-VP160 resulted in activation of only the target genes even when multiple sgRNAs were used.[26,52,56]

Transcriptomic analysis of animal cells expressing second-generation activators dCas9-VPR, SAM, and SunTag demonstrates the specificity of these PTAs even when higher target gene expression levels are achieved.[33,56] In plants, direct activation of *RPL23* by dCas9-TV was shown to be specific, but a few nontarget genes were indirectly induced because of RPL23 expression.[14] Similarly, activation of *FLOWERING WAGENINGEN (FWA)* by the SunTag system in the model plant Arabidopsis resulted in specific activation of *FWA* with only a few other upregulated genes.[60]

In Drosophila, RNA sequencing revealed mis-regulation of many genes besides the target genes, twist and snail, activated by dCas9-VPR. Because twist and snail are transcription factors themselves, some of the mis-regulated genes may represent direct and indirect targets and not result from off-target PTA activation.[57]

The apparent specificity of CRISPR-Cas activators may be due to the fact that they need to be targeted to a couple of hundred base pairs upstream of the TSS for maximum efficiency (Fig. 2e) (see Position Effects section). Most of the possible off-target binding sites simply will not produce a measurable phenotype.

When bound to a target site, the influence of CRISPR-Cas activators in the transcription of the surrounding genomic regions is limited. Analysis of 112 genes using a SunTag screen combined with singlecell RNA-sequencing (Perturb-Seq) showed that activation of a target gene does not affect the expression of neighboring genes unless they share promoter regions.[61] In addition, the dCas9-VP64 ribonucleoprotein complex appears to be sensitive to guide-target mismatches; it can tolerate only three mismatches at its binding positions, while off-target sites for Cas9 nucleases may contain up to five.[6,58] Even with these considerations, the careful design of sgRNAs that provide strong transcriptional activation with minimal offtarget sites will be the best way to provide a potent and specific induction of the desired genes.

1.6 Applications of CRISPR-Cas Activators

The advantages provided by dCas9-based PTAs have been leveraged in genetic screens and the creation of novel therapies. Most of the development of novel transcriptional activators, especially in mammals, has been carried out using transient transfection in cell lines. The ability to activate gene expression combined with the generation of genome-wide sgRNA libraries has allowed for the generation of novel gain-of-function (GOF) screens. In addition, *in vivo* applications using CRISPR-Cas have been developed in whole multicellular organisms.

1.6.1 - Genome-wide screenings using CRISPR-Cas activators

Among the early applications of dCas9-based PTAs is the development of genome-wide activation screens using sgRNA libraries (CRISPRa libraries) to identify genes whose overexpression will confer a phenotype easily scored in a high-throughput manner.[33,52] sgRNA enrichment in cells displaying the phenotype-of-interest is measured via high-throughput sequencing and used to identify the target GOF genes.[52]

Using the SAM system, Konermann et al. designed a screen to identify genes whose overexpression protects A375 malignant melanoma cells from cell cycle arrest and apoptosis induced by the BRAF inhibitor, PLX4720.[33] The lentiviral expression library designed for this screen, consisting of three sgRNA2.0s per every coding isoform (23,430 isoforms) of the human RefSeq database, was transformed into A375 cells expressing dCas9-VP64 and MCP-p65-HSF1. After selection in PLX-4720, enriched sgRNAs were sequenced and their corresponding target genes identified. These target genes, 13 of which were independently validated, were shown to correspond to known but also novel targets of PLX-4720.[33]

Another GOF genome-wide screen used a library composed of sgRNAs binding to 10 sites upstream of the TSS of 15,977 human genes.[52] After this library was transformed into K562 human cells expressing the SunTag system, one screen was used to determine enrichment of sgRNAs before and after 10 days of growth, and

another after exposure to a chimeric cholera/diphtheria fusion toxin (CTx-DTA).[52] Thus, while one screen identified genes affecting cell growth, the other identified genes modulating the response to the CTx-DTA toxin.[52] Among the overexpressed genes that cause growth suppression were tumor suppression genes, transcription factors involved in tissue development and differentiation, and mitotic genes.[52]

Similar screens using genome-wide CRISPRa libraries have been performed to identify overexpressed genes promoting neuronal differentiation, reprogramming of somatic cells to induced pluripotent stem cells, and conferring resistance to the anticancer drug imatinib.[62–64]

In contrast, other screens have made use of specific libraries to activate a specific subset of transcripts.[65,66] For instance, a CRISPRa library designed to activate all putative cell surface proteins was used to find extracellular receptors recognizing ligands of interest.[66] Another CRISPRa library targeting long-noncoding RNAs (lncRNAs) was used to find overexpressed lncRNAs conferring resistance to the PLX-4720 analogue, vemurafenib.[65] Interestingly, when some of these GOF genome-wide screens were paired with loss-of-function CRISPR interference (CRISPRi) screens, both provided complementary and comprehensive insights toward the phenotypes interrogated.[52]

CRISPRa/i screens allow for the systematic identification of individual genes associated with the phenotypes of interest. However, a biological phenotype often results from synergistic interactions of combinations of genes rather than the summed activity of individual genes. Thus, a genetic interaction (GI) between two genes will result in a deviation of the expected phenotypes resulting from simply adding their phenotypic effects.[67]

To measure interactions between the genes identified in CRISPRa screens, additional GI libraries have been generated and tested. To find GIs among 19 genes

identified in a screen for factors promoting neuronal differentiation, Liu and colleagues developed a combinatorial CRISPRa gene activation library. The lentiviral library consisted of a combination of paired sgRNAs, validated in the previous screen, each of which will activate a target gene to two levels (high or low). To allow comparisons between the effects of activating a single gene versus pairs, paired sgRNAs in which one will target a gene and the other will not were also included in the library.[62] This screen not only identified positive and negative interactions between genes promoting neuron formation, but also sheds light on the role that expression levels play in the intensity of the GI.[62] It also allowed for the discovery of gene pairs that can readily induce differentiation of fibroblasts into neurons.[62]

For a screen of GI modulating resistance to the drug imatinib, two orthogonal CRISPR-Cas9 systems, one producing gene knock outs (*Staphylococcus aureus* Cas9 nuclease) and the other activating gene expression (SunTag system using S. pyogenes dCas9), were used. The library design contained pairwise sgRNAs, one for SaCas9 nuclease and other for the SunTag system, so that one of the genes will be activated while the other knocked out. It also contained sgRNA pairs that will knock out or activate a single gene, while not perturbing the other.[63]

The sgRNAs in this library target 87 genes identified in a previous screen for imatinib resistance and 1327 genes involved in cancer-associated pathways combined in a total of 100,000 pairwise combinations.[63] As expected, a set of positive and negative GIs were seen. However, by evaluating the range of interactions resulting from combining gene activation and knock outs, this screen was able to determine the direction of the GI between pairs of genes, thereby allowing the inference of more complex interaction maps than those obtained with pairwise CRISPRa libraries.[63]

While initial GOF library screens were successful in identifying genes involved in the cellular processes analyzed, improvements in sgRNA design may produce libraries

with enhanced performance and result in more robust screens. In the first two CRISPRa libraries described, for each gene, multiple sgRNAs were designed to target a couple of hundred base pairs upstream of the TSS.[33,52]

However, to obtain libraries with improved sgRNA activity, a machine learning approach was used that created a predictive model incorporating nucleosome positioning, sequence features, and improved sgRNA design rules.[68] This model allowed the design of more active sgRNAs that in turn were used to construct an improved version (CRISPRa v2) of the library previously used by Gilbert and colleagues (CRISPR v1).[68] When tested, the CRISPRa v2 library was shown to identify 540 genes affecting cell growth, whereas a similar screen with CRISPR v1 identified only 283 genes.[68] Furthermore, the sgRNAs in the CRISPRa v2 library were shown to be more active than in CRISPR v1.

Another optimized library was created by using two approaches to modify an SAM-like system.[69] First, a novel guide, tracr-v14, was designed that includes two MS2 and two PP7 stem loops, thereby increasing the flexibility in the use of recruiting domains. Second, a narrow window (150–75 nucleotides upstream of the TSS) in which sgRNA activity was predicted to be highest was used for targeting.[69]

The resulting CRISPRa library, named Calabrese, was tested using a screen for resistance to a BRAF inhibitor, similar to the one used by Konermann and colleagues.[69] The screen performed with the Calabrese library was able to identify previously known and also novel genes that when activated confer resistance to vemurafenib. In addition, the number of genes identified using the Calabrese library was substantially larger than those identified in a similar screen performed by Konermann and colleagues using the SAM library.[69]

Thus, as suggested by these studies, refinement of the sgRNA design rules and accurate gene annotation, especially TSS and nucleosome positioning, are crucial for

the design and construction of CRISPRa libraries with significantly increased activation potential.

1.6.2 - Use of CRISPR-Cas activators in in vivo systems

CRISPR-Cas activators and their cognate sgRNAs have also allowed for the development of novel gene and cellular therapies. Some of these *in vivo* studies involve either injecting transfected cells or viral constructs containing the activators and sgRNAs into adult organisms or making use of transgenic animals to express the components of the PTA system in the desired organs or tissues (or a combination of both).

For instance, dCas9-VP64 was used in mice to identify genes that affect the sensitivity to the DNA-damage inducing chemotherapy drug, temozolomide (TMZ). B cell lymphoblastic leukemia cells (B-ALL) transfected with dCas9-VP64 and sgRNAs targeting the *Mgmt* (O6- methylguanine–DNA methyltransferase) gene were shown to be more resistant than control cells to TMZ treatments after lymphoma transplantation in adult mice.[70] Resistance of B-ALL cells to TMZ treatment required activation of Mgmt, a gene involved in the detoxification of TMZ-induced DNA damage. Using this system, a screen was designed to test for gene regulators of the DNA damage response that mediate sensitivity to TMZ.[70]

A GOF screen using the SunTag system was developed to study the effect of gene activation in a mouse liver injury and repopulation model. In this screen, the dCas9-GCN4 component of SunTag was expressed as a Cre-inducible transgene in mice. Then, to activate gene expression, the activator component (ScFv-VP64), Cre recombinase, and sgRNAs were delivered by injection before induction of liver injury. After liver repopulation, highthroughput sequencing determined which sgRNAs, and by extension which target genes, promote cell proliferation and the formation of hepatocellular carcinomas.[71] A similar approach using SPH, a SunTag-like system that

uses p65-HSF1 instead of VP64 as an AD, was used for the *in vivo* multiplexed activation of genes in the brain.[72]

Another GOF screen, this time for genes whose transcriptional activation induced heart failure in mice, was also recently developed. This system combined the expression of dCas9-VPR driven by the *Myh6* (myosin heavy chain) promoter in the heart of transgenic mice with the injection of adeno-associated virus (AAV) vectors containing sgRNAs to induce gene expression of target genes specifically in cardiomyocytes.[73]

An antitumor therapy called multiplexed activation of endogenous genes as an immunotherapy (MAEGI) was developed using PTAs. In MAEGI, the SAM system was used to induce the expression of endogenous genes in tumors, some of which encode tumor antigens that enhance antitumor immune responses.[74] To induce the multiplexed expression of tumor genes, a genome-scale or customized tumor-specific sgRNA2.0 library, including the MCPp65-HSF1 module, was created in AAV vectors. Intratumoral injection of AAV libraries in dCas9-VP64 expressing mice bearing orthotopic tumors led to increased tumor remission. Both libraries elicit a potent and specific antitumor immune response that showcases the therapeutic efficacy of MAEGI.[74]

In another study, the SAM system was used in rats for multiplexed activation of two osteogenesis-promoting genes, *Wnt10b* and *Forkhead c2* (*Foxc2*), in bone mesenchymal stem cells (BMSCs). These BMSCs, with activated *Wnt10b* and *Foxc2*, when implanted in gelatin scaffolds to calvarial bone defects were able to greatly enhance bone healing.[75]

Therapies to correct pathologies caused by haploinsufficiency, due to the loss-of-function of one gene copy, have been designed using CRISPR-Cas activators to increase expression of the remaining wild-type copy.[76] The proof of concept of this therapy was applied to either *Single-minded homologue 1* (*Sim1*) or Melanocortin 4

receptor (*Mc4r*) haploinsufficient mice (*Sim1+/-* and *Mc4r+/-*, respectively) that develop obesity phenotypes.

Expression of dCas9-VP64 and a single sgRNA targeting either the *Sim1* promoter or the *Sim1 candidate enhancer 2* in transgenic mice increased the expression of Sim1 to levels high enough to correct the obesity phenotype of *Sim1+/-*. Even though dCas9-VP64 was expressed constitutively, it only increased transcription in tissues in which *Sim1* was already expressed, suggesting a need for a tissue-specific transcription factor for *Sim1* expression.

As an alternative approach, dCas9-VP64 and sgRNAs, targeting *Sim1* or *Mc4r*, were delivered using AAV vectors injected directly to the tissues in which the target genes are active, the hypothalamus in this case. AAV-mediated delivery of the PTA system led to increased expression of the target genes and mice with long-lasting reduced weight, suggesting the feasibility of this approach to treat haploinsufficiency.[76]

In Drosophila, the first study of CRISPR-Cas activators in transgenic flies was carried out using dCas9- VP64 and dCas9-VPR activated by the Gal4-UAS system and two sgRNAs (sgRNA-wg) to target the *wingless* (*wg*) gene. Constitutive expression of dCas9-VP64 or dCas9-VPR was not toxic in transgenic flies. Furthermore, when dCas9-VPR was expressed together with the two sgRNA-wg, ectopic *Wg* production was observed accompanied by morphological abnormalities reminiscent of *Wg* overexpression phenotypes.[57] In contrast, dCas9- VP64 was not able to activate *Wg* expression, which is consistent with results obtained from cell lines.[57]

Ewen-Campen et al. also used dCas9-VPR and sgRNAs to activate several genes in transgenic Drosophila.[77] Out of 36 target genes, they observed activation of 27 (75%) although to a consistently lower expression than that obtained by expressing cDNAs of the same genes driven by the Gal4-UAS system.[77] Despite the lower activation levels, several of the predicted overexpression phenotypes were observed.[77]

Attempts to use the SAM system in Drosophila transgenic lines were motivated by previous results showing better performance than other activators in cell lines.[56,77] Expression of the SAM system in transgenic flies resulted in lethality that was avoided when its AD was replaced by ADs previously used in Drosophila.[77] However, none of these modified SAM systems performed better than dCas9-VPR in vivo. One of the reasons for the lethality of SAM was believed to be the high expression levels achieved in initial constructs, because even dCas9-VPR was shown to be toxic when expressed at higher levels.[77]

With these considerations, the SAM system was modified by expressing from a weaker promoter dCas9-VP64 in tandem with MCP-p65-HSF1, separated by a T2A self-cleaving peptide. This system, dubbed FlySAM, is not lethal or toxic when expressed in transgenic Drosophila and performs significantly better than dCas9-VPR at activating endogenous genes.[78] Phenotypes induced by FlySAM, even with the use of a single sgRNA2.0, are similar in severity to those produced by the Gal4-UAS overexpression systems.[78]

Inspired by the success of dCas9-VPR and FlySAM, Zirin et al. developed The Transgenic RNAi ProjectCRISPR overexpression (TRiP-OE) collection of Drosophila lines.[79] Each TRiP-OE stock expresses either sgRNAs or sgRNA2.0s for FlySAM, binding to the region upstream of the TSS of a target gene. Gene activation of the target gene is initiated by crossing the sgRNAs or sgRNA2.0s containing stocks with another harboring the dCas9-VPR or FlySAM activators, respectively.[79] There are also some stocks that express sgRNA2.0 together with FlySAM under the UAS control that are activated by crossing the lines with the Gal4-UAS activating system.[79]

In the nematode *Caenorhabditis elegans*, dCas9-VP160 was tested as an activator to induce expression of endogenous genes. Expression of dCas9-VP160 and single sgRNAs targeting dbl-1 (a transforming growth factor-b family gene) in transgenic

nematodes did not provide any target gene activation. Only when expressing the PTAs together with six sgRNAs targeting the *dbl-1* promoter was a modest increase in gene expression observed. The level of activation was, however, enough to induce the elongated body length phenotype expected from *dbl-1* overexpression.[80]

In plants, a few PTAs and target genes have been tested for activation in vivo. Transgenic plants expressing dCas9-VP64 and three sgRNAs targeting either *PRODUCTION OF ANTHOCYANIN PIGMENT 1* (*PAP1*) or the microRNA gene, *miR319*, were made in the model plant Arabidopsis thaliana.[81] Although these plants showed modest activation of the target genes (up to sevenfold increase), no phenotype was observed in any of the plants suggesting that higher expression levels may be needed to induce an overexpression phenotype.[81]

SAM and SAM-like systems have also been used to induce gene expression in transgenic plants. The SAM system with two sgRNAs2.0s was used in Arabidopsis to individually activate the expression of two endogenous genes, *PAP1* and *ARABIDOPSIS VACUOLAR H+-PYROPHOSPHATASE* (*AVP1*). SAM was capable of inducing only moderate gene expression in both target genes, but robust enough to display overexpression phenotypes.[82] An SAM-like system, in which the p65-HSF1 activator was replaced by VP64, was designed independently by two groups and used to generate *Arabidopsis* transgenic plants. One of these SAM-like systems, namely CRISPR-Act2.0, was used for the targeted overexpression of *PAP1* and *Fertilization-Independent Seed2* (*FIS2*), whereas the other SAM-like system was used to target the flowering-promoting gene *FLOWERING LOCUST T* (*FT*).[83,84]

Both SAM-like systems were able to induce gene expression of their target genes by one to two orders of magnitude. However, only the expected early flowering phenotype resulting from the activation of *FT* using the SAM-like system was reported.[83]

Other CRISPR-Cas systems have been used to generate transgenic plants. As shown in the dCas9-TV section, dCas9-TV was used to increase the expression levels of receptor *RLP23* leading to plants with enhanced immune response.[14] The SunTag system was used in Arabidopsis to activate the expression of three genes, *FWA*, *CLAVATA3* (*CLV3*), and *APETALA3* (*AP3*). SunTag-mediated activation of these genes reached several hundred-fold transcript levels compared with wild-type, and phenotypes associated with the overexpression of *FWA* and *CLV3* were observed in transgenic plants.[60]

Thus far, to our knowledge, there has not been any reports on the toxicity or lethality associated with the expression of CRISPR-Cas activators in plants, even when expressed from strong constitutive promoters. However, protein instability from the expression of dCas9-TV designs with an increased number of AD repeats, and toxic effects specifically associated with the activation of *PAP1* in Arabidopsis have been observed.[14,82] Thus, in comparison with Drosophila, plants appear to be more resilient to the constitutive expression of activators.

1.7 Future Remarks

The development of CRISPR-Cas-based PTAs capable of robust and specific gene induction and with the ease of multiplexing provides the means to design genetic screens to interrogate biological systems, especially transcriptional networks. Novel biotechnologies based on the rewiring of transcriptional programs have emerged from these developments, such as the metabolic engineering of organisms.[34,54,85,86] PTAs may also lead to the development of therapies based on the corrective activation of genes or pathways that fail to express during disease. In insects, PTAs have been used to engineer synthetic speciation events, an approach that could improve the biocontainment of transgenes and lead to novel biocontrol approaches for pests and disease vectors.[87]

Continuous development of new activator architectures that allow even further activation levels will pave the way for new applications. Furthermore, combination of these activators with other CRISPR-Cas effectors such as repressors or chromatin modulators will exponentially increase the range of applications in which they could be used. However, more research is needed in the performance of these activators in in vivo systems, especially in vertebrates. In addition, deep knowledge of the transcriptional networks involved in the processes of interest is required to efficiently deploy PTAs to obtain the desired outcomes.

1.8 – Figure Legends

Figure 1 - Diagram of different CRISPR-Cas transcriptional systems. For each PTA described, a schematic representation of the CRISPR-Cas protein in complex with its sgRNA, target DNA sequence, and additional protein modules, when present, is shown in the upper panels. Expression constructs encoding the protein components of the activator systems are shown in the lower panels. Each construct is driven by the human cytomegalovirus promoter and the HSV thymidine kinase polyadenylation signal sequence as a terminator, unless indicated otherwise. Genetic elements are drawn to scale. Cas proteins are shown in different shades of gray, and the sgRNA and crRNA are indicated by blue lines inside the Cas proteins. ADs are shown in different colored shapes. (a) dCas9-VP64. (b) dCas9-VPR. (c) dCas9-TV. (d) scRNA. The MS2 loops appended to the sgRNA to form the (scaffold sgRNA) scRNA are indicated by purple lines. (e) SAM. The MS2 loops in the sgRNA2.0 are indicated by purple lines. (f) TREE. (g) SunTag. (h) split ddCpf1. The N-terminal and C-terminal fragments of ddCpf1 are indicated as N-ddCpf1 and C-ddCpf1, respectively. (i) Sth Cascade. The different domains that form the DNA binding Cascade complex are CasA, A; CasB, B; CasC, C; CasD, D; and CasE, E. Cascade expression constructs are driven by the ZmUBI promoter and the PinII terminator. (j) Schematic representation of different sgRNA expression cassettes used in the CRISPR-Cas transcriptional activators shown. Unless indicated otherwise, the human U6 RNA

polymerase III promoter is used in all constructs. The terminator in all sgRNA constructs is a poly T-stretch of 6-8 nucleotides. sgRNA scaffold is represented by an open box and the ps in blue. AD, activation domain; HSV, herpes simplex virus; PinII, potato proteinase inhibitor; ps, protospacer; PTA, programmable transcriptional activator; SAM, synergistic activation mediator; scRNA, scaffold RNA; sgRNA, small guide RNA; TREE,

three-component repurposed technology for enhanced expression; ZmUBI, Zea mays ubiquitin.

Figure 2 - Features of transcriptional stimulation mediated by CRISPR transcriptional activators. (a) Synergy. When two dCas9-VP64 copies are recruited, they act concertedly to yield stronger transcriptional activation than a single copy (left panel). In the right panel, dCas9-VP64 can induce the production of VEGFA protein. However, VEGFA protein production is synergistically increased when three or four sgRNAs are coexpressed with the dCas9-VP64 activator.30 (b) Position effects. Binding of dCas9-VP64 to the region upstream to the TSS induces robust gene activation (right panel). Induction of expression in 12 human genes using the SAM activator is the strongest when using sgRNAs located between 200 and 0 bp of the TSS.33 (right panel) (c) Gene to gene variability. The relative change in gene expression induced by dCas9-VP64 will be larger for weakly induced genes than for genes with higher basal expression (left panel). There is a negative correlation between the basal expression of a gene and the fold-activation attained by dCas9-VPR.13 (right panel). (d) Multiplexing. dCas9-VP64 can be used for the targeted activation of multiple genes by using pooled sgRNAs binding to the promoter of different genes. (e) Specificity. Despite having more than one binding site, dCas9-VP64 specifically activates the expression of its target genes with little to no off-target effects. This is, in part, due to the position effect requiring dCas9-VP64 to target the region upstream of the TSS for maximum efficiency. The TSS is indicated by a bent arrow. The thickness of the arrow represents the strength of gene expression. The dCas9-VP64 activator consists of the dCas9 protein in gray, the VP64 activation domain in green, and the colored line inside the dCas9 protein is the sgRNA. VEGFA, vascular endothelial growth factor A; TSS, transcription start site.

Table 1 – Summary of CRISPR-Cas activation systems

1.9 – Tables

Table 1 – Summary of CRISPR-Cas activation systems

PTA system	CRISPR Cas system[a]	No. of elements[b]	ADs/copy number recruited[c]	Organisms	Addgene catalog number	References
First-generation						
dCas9-AD	Class 2/Cas9, Class 2/Cpf1 (Cas12a)	2	VP64/1X; VP640/1X; p65/1X; EDLL/1X; TAL/1X; CBF1/1X	Mammals, fish, worms, insects, yeast, plants	48214, 48218–48228, 48236–48240, 49013–49016, 46912–46923, 50918, 50920, 47753, 47754, 47314–47321, 47106–47108, 69303, 132334	6,13,20,26–32,40,80,81
Second-generation						
dCas9-VPR	Class 2/Cas9, Class 2/Cpf1 (Cas12a)	2	VP64-p65-Rta/1X	Mammals, insects, yeast, plants	63798–63802, 64046, 104567	13,39,40,88
dCas9-TV	Class 2/Cas9, Class 2/Cpf1 (Cas12a)	2	TAL/6X, VP64/2X	Plants, mammals	n.a.	14
SAM	Class 2/Cas9	3	VP64/1X, p65-HSF1/4X; VP64 5X	Mammals, insects, plants	78901, 78902, 78905, 61422, 61427, 73795, 99884–99897, 99905–99907, 100044, 122835–122839, 122856, 122857, 122860, 122861	33,56,83,84,88
scRNA	Class 2/Cas9	3	VP64/4X; EDLL/1X, VPR/4X	Mammals, yeast, plants	47314–47321, 78906, 62277, 62279–62283, 62313–62322, 62325, 62327, 62328, 62330–62342, 62344, 66564, 66565	6,34,56,88
SunTag	Class 2/Cas9, Class 2/Cpf1 (Cas12a)	3	VP64/10X; p65-HSF1/10X	Mammals, insects, plants	60903, 60904, 78899, 78900, 119672, 120249–120252	35,41,56,60,72
TREE	Class 2/Cas9	4	VP64/1X, HSF1-p65/16-32X; VP64/1X, VP64-p65-Rta/16-32X	Mammals	na	36
Split ddCpf1	Class 2/Cpf1 (Cas12a)	3	HSF1-p65/4X	Mammals	na	42
Sth Cascade	Class 1/CRISPR-Cascade	6	CBF1/3X	Plants	132334–132340, 132342–132353	20

[a]The number of elements takes into account the sgRNA or sgRNA2.0 component.
[b]A semicolon (;) separates different activator sets used in the same system.
[c]Copy number refers to the theoretical maximum of ADs recruited.
AD, activation domain; na, not available; PTA, programmable transcriptional activator; SAM, synergistic activation mediator; scRNA, scaffold RNA; TREE, Three-component repurposed technology for enhanced expression.

1.10 - Figures

Figure 1 - Diagram of different CRISPR-Cas transcriptional systems.

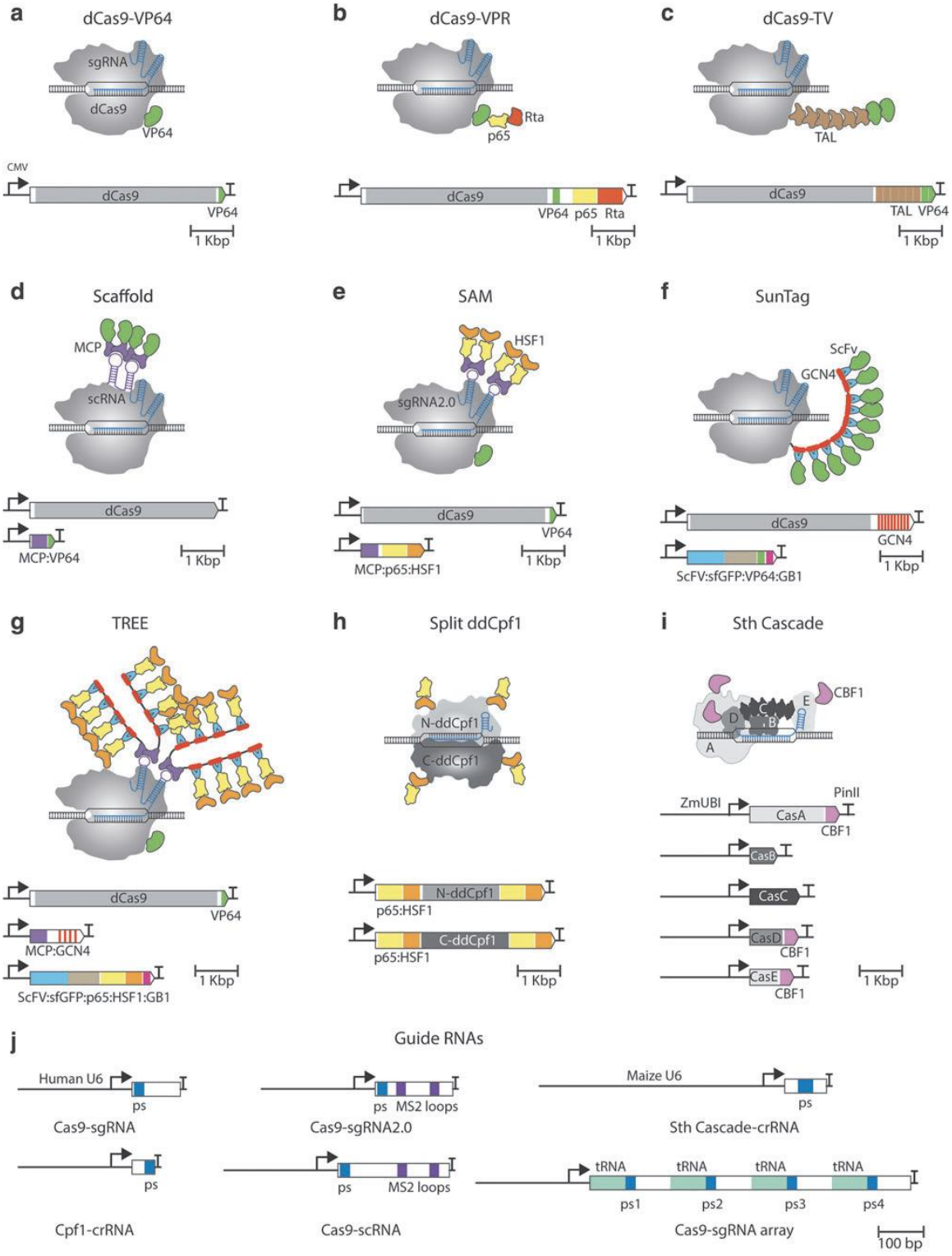
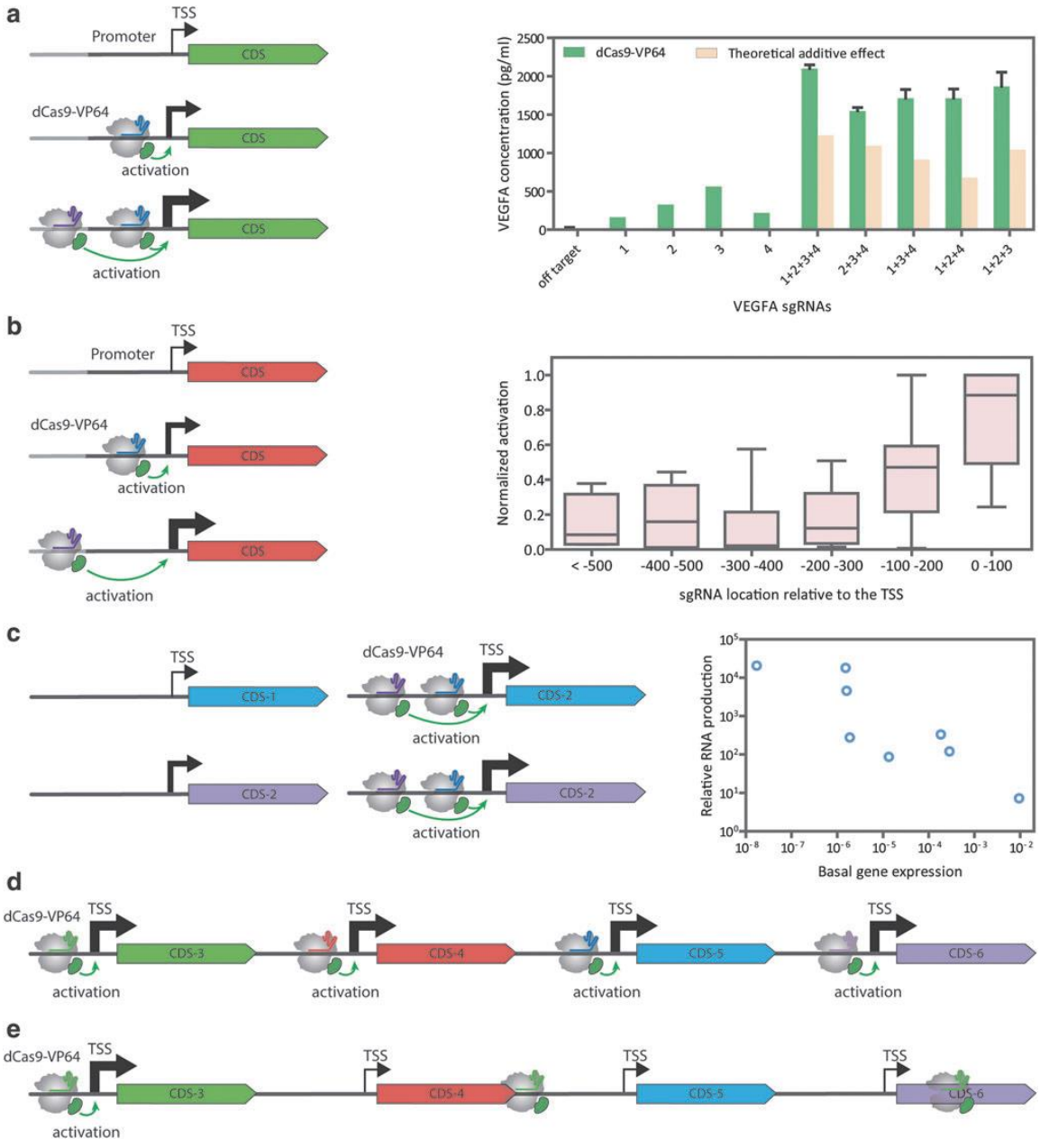


Figure 2 - Features of transcriptional stimulation mediated by CRISPR transcriptional activators



Chapter 2

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], Daniel F Voytas[1,3,4], and Michael J Smanski[2,3,4]

[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

2.1 Abstract

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 transactivation 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 transactivation 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.

2.2 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.[88,89] Studies correlating differential gene expression with phenotype allows for predicting which genes are most important for proper development and responding to the environment.[90,91] 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, 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)[92] PTAs can drive the over- or ectopic expression of endogenous genes when designed to bind to a promoter region.[30,31] 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[93]. dCas9-based PTAs can target many promoters in parallel by co-expressing multiple sgRNAs[72]. PTAs have been used to explore gene function and engineer barriers to sexual reproduction in insects[94].

The fist 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[84]. An additional improvement to PTAs came with dCas9 translational fusion to an AD termed VPR[13]. 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.

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[14]. 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[35]. Using 10 copies of GCN4 fused to dCas9, the SunTag system can activate target genes in *Arabidopsis thaliana* to produce visible phenotypes[60]. 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 report the discovery and performance of several plant-evolved ADs for programmable gene expression in monocot and dicot plants.

2.3 - 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[95,96].

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[97] 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 ex- pressed 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 (Supplementary Fig 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 di- verse 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[10]. 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 identity across coding sequences from which ADs were derived, as illustrated by the low-confidence bootstrap values in a Maximum Likelihood tree comparing these sequences (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 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.

2.4 - A set of seven 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[98,99]. On the reporter plasmid, we designed a synthetic promoter comprising six copies of the *Lac* operator upstream of the minimal 35S CaMV core promoter upstream of Firefly luciferase[100]. 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 (Fig. 1c).

Alternative activation domains were expressed as scFv-AD fusions driven by a 35S promoter, allowing for strong ex- pression 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* (Fig 1a). The library was first tested in *Setaria viridis* protoplasts. We isolated and transformed 100,000 cells with 2 ug 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 dataset showed a weak correlation (R=0.15, Supplemental Fig 2), when fitting the second data set to a linear model we noticed a strong correlation between Firefly and Renilla luciferase (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 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 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 (Fig 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 luciferase platform in *A. thaliana* protoplasts with a smaller library of ADs and saw a weak correlation for a linear regression model (Fig 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 Setaria datasets (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 (Fig 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 TADs 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 ex- pression 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.

2.5 - DREB2, DOF1, and AvrXa10 plant-derived activation domains outperform VP64 across endogenous loci.

We next set out to test 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 (Fig 2a-b, Supplementary Figures 6-8). All sgRNAs used in this study are previously published[60,83,84], except for the *Wuschel* 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 house- keeping gene (PPa2) were amplified. We report expression relative to PP2A in Figure 2b for FT, PAP1, and WUS. By graphing all

datapoints 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.[93]

The sgRNAs bind the core promoter regions, located 0-300 base pairs upstream of an annotated transcription start site (TSS) for the gene (Figure 2a). The Pap1 sgRNA used in this study interestingly 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 2a-c, 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), but this is due to FT starting at much lower basal expression levels. The maximum absolute expression attained is comparable. WUS, on the other hand, could only be overexpressed to 10% the level of PP2A (Figure 2b).

To then view the data in aggregate, we compared each AD's gene activation dataset to VP64. After setting VP64 activation values for each gene tested equal to 1, we combined datasets for all sgRNAs across four target genes (n=13). The same 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 next set out to test 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) (cite). 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 20bp region 39bp upstream of the transcription start site for the *A. thaliana* gene *Lec1*. In addition, we cloned an sgRNA targeting a 20 bp region 139bp upstream of the transcription start site for 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.

2.6 - AvrXa10, DOF1, and DREB2 activation domains promote early flowering in transgenic plants.

Next, we tested whether our plant-derived ADs would produce a phenotype upon gene activation in stable transgenic plants. We generated T-DNA vectors expressing our plant-derived ADs using a new PTA architecture named the MoonTag system.[101] The MoonTag system utilizes an epitope-nanobody interaction to recruit the AD to the dCas9

molecule.[102] 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.[103,104] For comparison, we used a validated Moontag-VP64 transgenic line as the positive control[101] 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 3c,d).

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.[105,106] 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 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, and AvrXa10 showed the strongest molecular and phenotypic gene activation within the dicot species *Arabidopsis thaliana*.

2.7 - 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.[107] To illustrate this concept in plants, we utilized the FT gene activation platform. Of the two sgRNAs tested, FTgB could drive stronger gene activation than FTgA at the core promoter in our protoplast assays (Fig 2, 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.[108] 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 gRNA 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-gRNA 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 gRNA, as the gRNAs targeting sequences outside of the enhancer regions did not significantly boost expression levels.

2.8 - 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 isolation and transformation serves as a valuable tool for testing large numbers of genetic parts[96]

While performing these assays, we observed what appeared to be local off-target activation of our dual luciferase reporter. 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 up to 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. This phenomenon 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.[52]

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 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.[10] 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 phenotyping was carried out on the T2 generation for quantification of the early flowering phenotype. 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[52] 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.[109] PTA-mediated enhancer activation retains the original genomic architecture of the enhancer- promoter pair, while traditional luciferase assays deploy

enhancer sequences in non-native genomic architectures. Given our understanding of enhancers functioning through a DNA looping mechanism, retaining this genetic architecture may be important in identification of enhancers.

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.[110] 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.

Expanding the definition of a gene to include enhancers has upside for plant engineering in a variety of contexts. First, integration of transgenes by *A. tumefaciens* is a random process. If a transgene lands between a core promoter and distal enhancer, this may prevent the enhancer from acting on the core promoter. Thus, identifying promoter-enhancer pairs can aid in selecting transgenic lines with undisturbed gene expression due to transgene integration site. In addition, these enhancer regions could be targets for new genetic variation, either natural variation or a target for gene editing to introduce new variation, which may lead to changes in gene expression and phenotype. For example, CRISPR-Cas9 mutagenesis generating large deletions in the space

between a distal enhancer and the core promoter(s) it regulates may result in changes to gene expression, potentially leading to new promoter-enhancer interactions.

This study has limitations that should be addressed with future work. The PTAs we describe here work great in research settings, but they have components derived from human or plant pathogens that would trigger 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.[111] 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.[112] 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.

2.9 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 CTD 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.

Protoplast isolation and transformation. *S. viridis* and *A. thaliana* protoplasts were isolated and transformed as stated in previous protocols.[95,96] ME034 *S. viridis* plants are

grown in a growth chamber set to 31C and 21C 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. Zea mays plants were grown in a growth chamber at 25C 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 midi prepped 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[98]. 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 only a transcription start site. 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 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 to 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. 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 55C and yield amplicon lengths be- tween 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.

Plant transformation and genotyping. To generate transgenic *A. thaliana* plants we performed floral dip protocols as previously described.[103] TDNA plasmids were

constructed containing an antibiotic resistance gene along with the pFAST Oleosin-RFP transgene. *A. thaliana* ecotype Columbia-0 was grown to maturity and floral dipped with Agrobacterium strain GV3101 transformed with the described TDNAs. 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.

2.10 - Figure Legends

Figure 1. Screening a library of plant-derived activation domains with the dual luciferase protoplast assay. (a) Cartoon images of dCas9-based Programmable Transcription Activators (PTAs) including the direct fusion (top) and SunTag scaffolding system (bottom). Genetic construct design for corresponding plasmids transformed into S. viridis or A. thaliana protoplasts are shown below each cartoon image. (b) Design of a single-plasmid dual luciferase reporter construct (top), and strong correlation of the PTA-inducible Renilla luciferase activity vs 'constitutive' Firefly luciferase activity (bottom) (c)Same date as in (b), but with the transformation control Firefly luciferase split to a separate plasmid (d) Phylogeneic tree showing the diversity of sequences tested in this study. The multiple sequence alignment and phylogenetic tree were generated in MEGA using full length amino acid sequences. A Maximum Liklihood tree was generated using 500 bootstraps, with values >0.05 displayed above a given node. Strong activation domains from our assays are highlighted in green on the tree, while the positive control VP16 is highlighted in red. (e) A rank order of the strongest ADs is show for S. viridis 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.

Figure 2. Endogenous gene activation in A. thaliana protoplasts. (a) Guide RNA design for a given gene 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. (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 with TALE-SunTag. 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.

Figure 3. FT gene activation in transgenic plants. (a) Genetic construct design for T-DNAs floral dipped into transgenic A. thaliana plants. (b) The FT core promoter was targeted with two sgRNAs each binding within 200bp of the transcription start site. (c) Image of T2 transgenic plants. The wildtype Col-0 plant is shown on the left, the positive control VP64 line is shown in the middle, and the AvrXa10 plant for line 2 is shown on the right. (f) Molecular quantification of FT gene activation from 5d old T2 seedlings. (g) Bolting time quantification in T2 transgenic plants. (h) 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.

Figure 4. PTA-mediated enhancer activation at the Arabidopsis thaliana FT locus. (a) Guide RNAs were designed to target each of the indicated regions with varying distances to the known FT TSS. The FT Core promoter (-197bp), Block C (-5.3kb),Block B (-1.8kb), and Block E (+3.8kb) enhancers were targeted along with intervening sequences equidistant to a given enhancer and its nearest transcriptional regulatory element. All indicated distances for gRNAs are relative to the FT transcription start site. (b-d) FT gene activation when targeting Block C (b), Block B (c) or Block E (d) with different combinations of gRNAs. RNA was extracted 24h post transformation, and FT gene activation was quantified via RT-qPCR. Parametric t-tests were performed comparing each AD with the negative control where no gRNA is delivered. * corresponds

to p <= 0.05, ** corresponds to p <= 0.01, *** corresponds to p <=0.001, and **** corresponds to p<= 0.0001.

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.

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.

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.

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.

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.

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.

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.

Supplemental Figure 8. *A. thaliana* protoplast Clv3 gene activation with gRNA1. 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.

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.

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

2.11 Tables

Supplementary Table S1. Activation Domains

Name	Domain Source	AD Seq	Citation	Uniprot
Dof1	Plant	175-238 (QPGTEDAEAVALGLGLSDFPSAGKAVLDDEDSFVWPAASFDMGACWAGAGFADPDPACIFLNLP)	10.1093/pcp/pce105	http://www.uniprot.org/uniprot/P38564
DREB1/2	Plant	DREB1 (108-216 - GRSACLNFADSAWRLRIPESTCAKDIQKAAAEAALAFQDEMCDATTDHGFDMEETLVEAIYTAEQSENAFYMHDEAMFEMPSLLANMAEGMLLPLPSVQWNHNHEVDGDDDDVSLWSY)	10.1111/pbi.12057	http://www.uniprot.org/uniprot/Q9M0L0
		DREB2 (254-335 - SSDMFDVDELLRDLNGDDVFAGLNQDRYPGNSVANGSYRPESQQSGFDPLQSLNYGIPPFQLEGKDGNGFFDDLSYLDLEN)		http://www.uniprot.org/uniprot/O82132
IPN2	Plant	179-358 (AATNLKGIGPQTIPDMGIMKEFGSPLGFSFQDLDLYGGGGGDQLELQQNMEKPPLDGFMPMNHENLCLGKKRPNPYSGNNGKSPLMWSDDLRLQDLGSCLQDDPFKGDHHHQIQIAPPSLDRGTEMDPMSEIYDSKPEEKKFDASMKLERPSPRRAPLGERMSPMITTGTMAQGRSSPFG)	10.1111/nph.12593	https://www.uniprot.org/uniprot/E5L8F7
PTI4	Plant	170-234 (EPEPVRVTAKRRASPEPASSSGNGSMKRRRKAVQKCDGEMASRSSVMQVGCQIEQLTGVHQLLVI)	10.1111/pbi.12057	http://www.uniprot.org/uniprot/O04680
AtARF8	Plant	350-702 (VSLWEIEPLTTFPMYPSLFPLRLKRPWHAGTSSLPDGRGDLGSGLTWLRGGGGEQQGLLPLNYPSVGLFPWMQQRLDLSQMGTDNNQQYQAMLAAGLQNIGGGDPLRQQFVQLQEPHHQYLQQSASHNSDLMLQQQQQQQASRHLMHAQTQIMSENLPQQNMRQEVSNQPAGQQQQLQQPDQNAYLNAFKMQNGHLQQWQQQSEMPSPSFMKSDFTDSSNKFATTASPASGDGNLLNFSITGQSVLPEQLTTEGWSPKASNTFSEPLSLPQAYPGKSLALEPGNPQNPSLFGVDPDSGLFLPSTVPRFASSSGDAEASPMSLTDSGFQNSLYSCMQDTTHELLHGAGQINS)	10.1105/tpc.008417	https://www.uniprot.org/uniprot/Q9FGV1
SNAC1	Plant	172-314 (WEKMQQGKEVKEEASDMVTSQSHSHTHSWGETRTPESEIVDNDPFPELDSFPAFQPAPPPATAMMVPKKESMDDATAAAAAAATIPRNNSSLFVDLSYDDIQGMYSGLDMLPPGDDFYSSLFASPRVKGTTPRAGAGMGMVPF)	10.1073/pnas.0604882103	http://www.uniprot.org/uniprot/A2XNB9

Opaque-2	Plant	41-227 (IDVAAAGHGDGDMMDQQHATEWTFERLLEEEALTTSTPPPVVVVPNSCCSGALNADRPPVMEEAVTMAPAAVSSAVVGDPMEYNAILRRKLEEDLEAFKMWRADSSVVTSDQRSQGSNNHTGGSSIRNNPVQNKLMNGEDPINNNHAQTAGLGVRLATSSSSRDPSPSDEDMDGEVEILGFKM)	PMC146487	http://www.uniprot.org/uniprot/P12959
AvrXa10	Bacterial	TVMWEQDAAPFAGAADDFPAFNEEELAWLMELLPQSGSVGGTI	10.1094/MPMI.1998.11.8.824	https://www.uniprot.org/uniprot/Q56830
AL2/C2/TrAP	Viral	83-129 (PLHQHQDIPLTNQVQPQPEESIGSPQGISQLPSMDDIDDSFWENLFK)	10.1006/viro.1999.9925	http://www.uniprot.org/uniprot/P03562
AtARR2	Plant	279-664 (QHQGNMNHSFMTGQDQSFGPLSSLNGFDLQSLAVTGQLPPQSLAQLQAAGLGRPTLAKPGMSVSPLVDQRSIFNFENPKIRFGDGHGQTMNNGNLLHGVPTGSHMRLRPGQNVQSSGMMLPVADQLPRGGPSMLPSLGQQPILSSSVSRRSDLTGALAVRNSIPETNSRVLPTTHSVFNNFPADLPRSSFPLASAPGISVPVSVSYQEEVNSSDAKGGSSAATAGFGNPSYDIFNDFPQHQQHNKNISNKLNDWDLRNMGLVFSSNQDAATATATAAFSTSEAYSSSSTQRKRRETDATVVGEHGQNLQSPSRNLYHLNHVFMDGGSVRVKSERVAETVTCPPANTLFHEQYNQEDLMSAFLKQEGIPSVDNEFEFDGYSIDNIQV)	10.1111/j.1365-313X.2000.00909.x	https://www.uniprot.org/uniprot/Q9ZWJ9
RF2a	Plant	56-108 [RTBV promoter specific activation] AND/OR 283-357 [binds TBP] (EILSLPEDLDLCAAGGGDGPSLSDENDEELFSMFLDVEKLNSTCGASSEAEAE) (GGMMMNFGGMPHQFGGNQQMFQNNQAMQSMLAAHQLQQLQLHPQAQQQQVLHPQHQQQQPLHPLQAQQLQQAARDLKMKSPMGGQSQWGDGKSGSSGN)	10.1074/jbc.M304862200	http://www.uniprot.org/uniprot/Q69IL4
OsCBT	Plant	542-755 (PVTESLLELVLRNRLQEWLVEMVMEGHKSTGRDDLGQGAIHLCSFLGYTWAIRLFSLSGFSLDFRDSSGWTALHWAAYHGRERMVATLLSAGANPSLVTDPTPESPAGLTAADLAARQGYDGLAAYLAEKGLTAHFEAMSLSKDTEQSPSKTRLTKLQSEKFEHLSEQELCLKESLAAYRNAADAASNIQAALRERTLKLQTKAIQLANPEIEASEIV)	10.1074/jbc.M504616200	https://www.uniprot.org/uniprot/Q7XHR2
AP1	Plant	193-256 (HNMPPPLPPQQHQIQHPYMLSHQPSPFLNMGGLYQEDDPMAMRRNDLELTLEPVYNCNLGCFAA)	10.1023/A:1006273127067	https://www.uniprot.org/uniprot/P35631

MYC2	Plant	148-188 (GGVAPSDDAVDEEVTDTEWFFL VSMTQSFACGAGLAGKAFA)	10.1016/j.celrep.2017.04.057	https://www.uniprot.org/uniprot/Q39204
ZmVP1	Plant	1-120 (MEASSGSSPPHSQENPPEHGG DMGGAPAEEIGGEAADDFMFAE DTFPSLPDFPCLSSPSSSTFSSNS SSNSSSAYTNTAGRAGGEPSEPA SAGEGFDALDDIDQLLDFASLSM PWDSEPF)	10.1016/0092-8674(91)90436-3	https://www.uniprot.org/uniprot/P26307
OsGRF1	Plant	221-396 (PPPSYYSMDHKEYAYGHATKEV HGEHAFFSDGTEREHHHAAAGH GQWQFKQLGMEPKQSTTPLFPG AGYGHTAASPYAIDLSKEDDDEK ERRQQQQQQQHCFLLGADLRL EKPAGHDHAAAAQKPLRHFFDE WPHEKNSKGSWMGLEGETQLS MSIPMAANDLPITTTSRYHNDE)	10.1093/pcp/pch098	https://www.uniprot.org/uniprot/A2XA73
AtHSFA2	Plant	230-341 (KEKKSLFGLDVGRKRRLTSTPSL GTMEENLLHDQEFDRMKDDMEM LFAAAIDDEANNSMPTKEEQCLE AMNVMMRDGNLEAALDVKVEDL VGSPLDWDSQDLHD)	10.1111/j.1365-313X.2004.02111.x	https://www.uniprot.org/uniprot/O80982
AtHSFA6b	Plant	252-391 (KEKRKEIEEAISKKRQRPIDQGKR NVEDYGDESGYGNDVAASSSALI GMSQEYTYGNMSEFEMSELDKL AMHIQGLGDNSSAREEVLNVEKG NDEEEVEDQQQGYHKENNEIYG EGFWEDLLNEGQNFDFEGDQEN VDVGSSSHTN)	10.1111/j.1365-313X.2004.02111.x	https://www.uniprot.org/uniprot/Q9LUH8
EFS	Plant	820-1230 (KTTGALLDADIGKTSATYGTISSD VTHGEMVVDVTIEDSYSTESAWV RCDDCFKWRRIPASVVGSIDESS RWICMNNSDKRFADCSKSQEMS NEEINEELGIGQDEADAYDCDAA KRGKEKEQKSKRLTGKQKACFKA IKTNQFLHRNRKSQTIDEIMVCHC KPSPDGRLGCGEECLNRMLNIEC LQGTCPAGDLCSNQQFQKRKYV KFERFQSGKKGYGLRLLEDVREG QFLIEYVGEVLDMQSYETRQKEY AFKGQKHFYFMTLNGNEVIDAGA KGNLGRFINHSCEPNCRTEKWM VNGEICVGIFSMQDLKKGQELTFD YNYVRVFGAAAKKCYCGSSHCR GYIGGDPLNGDVIIQSDSDEEYPE LVILDDDESGEGILGATSRTFTDD ADEQMPQSFEKVNGYKDLAPDN TQTQ)	10.1105/tpc.109.070060	https://www.uniprot.org/uniprot/F4I6Z9
AFT1	Plant	Full length (MAATLGRDQYVYMAKLAEQAER YEEMVQFMEQLVTGATPAEELTV EERNLLSVAYKNVIGSLRAAWRIV SSIEQKEESRKNDEHVSLVKDYR SKVESELSSVCSGILKLLDSHLIPS AGASESKVFYLKMKGDYHRYMA EFKSGDERKTAAEDTMLAYKAAQ DIAAADMAPTHPIRLGLALNFSVF	10.1016/S0014-5793(98)01739-6	https://www.uniprot.org/uniprot/P48349

		YYEILNSSDKACNMAKQAFEEAIA ELDTLGEESYKDSTLIMQLLRDNL TLWTSDMQEQMDEA)		
EIN3	Plant	310-620 (SLSGGSCSLLMNDCSQYDVEGF EKESHYEVEELKPEKVMNSSNFG MVAKMHDFPVKEEVPAGNSEFM RKRKPNRDLNTIMDRTVFTCENL GCAHSEISRGFLDRNSRDNHQLA CPHRDSRLPYGAAPSRFHVNEVK PVVGFPQPRPVNSVAQPIDLTGIV PEDGQKMISELMSMYDRNVQSN QTSMVMENQSVSLLQPTVHNHQ EHLQFPGNMVEGSFFEDLNIPNR ANNNNSSNNQTFFQGNNNNNNV FKFDTADHNNFEAAHNNNNNSS GNRFQLVFDSTPFDMASFDYRDD MSMPGVVGTMDGMQQKQQDVSI WF)	10.1016/j.jmb.2005.02.065	https://www.uniprot.org/uniprot/O24606
WRKY50	Plant	1-106 (MNDADTNLGSSFSDDTHSVFEF PELDLSDEWMDDDLVSAVSGMN QSYGYQTSDVAGALFSGSSSCFS HPESPSTKTYVAATATASADNQN KKEKKKIKGRVAFKTR)	10.3389/fpls.2018.00930	https://www.uniprot.org/uniprot/Q8VWQ5
ATXR7	Plant	1250-1423 (DGADVLKMSQLKARKKHLRFQQ SKIHDWGLVALEPIEAEDFVIEYV GELIRSSISEIRERQYEKMGIGSSY LFRLDDGYVLDATKRGGIARFINH SCEPNCYTKIISVEGKKKIFIYAKR HIDAGEEISYNYKFPLEDDKIPCN CGAPNVYCFCEQVPWIAKLKRRT WFSRRN)	10.1105/tpc.109.070060	https://www.uniprot.org/uniprot/F4K1J4
ASF1a	Plant	Full length - (MSAIKITNVAVLHNPAPFVSPFQF EISYECLNSLKDDLEWKLIYVGSA EDETYDQLLESVLVGPVNVGNYR FVFQADPPDPSKIQEEDIIGVTVLL LTCSYMGQEFLRVGYYVNNDYE DEQLKEEPPTKVLIDKVQRNILSD KPRVTKFPIDFHPEEEQTAATAAP PEQSDEQQPNVNGEAQVLPDQS VEPKPEES)	10.1111/pce.12299	https://www.uniprot.org/uniprot/Q9C9M6
HAG1	Plant	150-550 (LESSDGGKDGGSSVVGTGVSGT VGGSSISGLVPKDESVKVLAENF QTSGAYIAREEALKREEQAGRLK FVCYSNDSIDEHMMCLIGLKNIFA RQLPNMPKEYIVRLLMDRKHKSV MVLRGNLVVGGITYRPYHSQKFG EIAFCAITADEQVKGYGTRLMNHL KQHARDVDGLTHFLTYADNNAVG YFVKQGFTKEIYLEKDVWHGFIKD YDGGLLMECKIDPKLPYTDLSSMI RQQRKAIDERIRELSNCQNVYPKI EFLKNEAGIPRKIIKVEEIRGLREA	10.1093/nar/29.7.1524	https://www.uniprot.org/uniprot/Q9AR19

		GWTPDQWGHTRFKLFNGSADMV TNQKQLNALMRALLKTMQDHADA WPFKEPVDSRDVPDYYDIIKDPID LKVIAKRVESEQYYVTLDMFVADA RRMFNNCRTYNSPDTIYYKCATR LETHFHSKVQAGLQSGAKSQ)		
GT-3A	Plant	110-323 (RYKACETTEPDAIRQQFPFYNEI QSIFEARMQRMLWSEATEPSTSS KRKHHQFSSDDEEEEVDEPNQDI NEELLSLVETQKRETEVITTSTST NPRKRAKKGKGVASGTKAETAG NTLKDILEEFMRQTVKMEKEWRD AWEMKEIEREKREKEWRRRMAE LEEERAATERRWMEREEERRLR EEARAQKRDSLIDALLNRLNRDH NDDHHNQGF)	10.1016/S0014-5793(04)00222-4	https://www.uniprot.org/uniprot/Q9SDW0
CVBp12	Virus	LCVDIYKRAFPRSVNKGRSSYAG GSGGVPGISYNSKVRDYILWGVT EVIP	10.1105/tpc.112.106476	https://www.uniprot.org/uniprot/P37992
ASF1b	Plant	Full length - (MSSINITNVTVLDNPAPFVNPFQF EISYECLTSLKDDLEWKLIYVGSA EDETYDQVLESVLVGPVNVGNYR FVLQADSPDPLKIREEDIIGVTVLL LTCSYMDQEFIRVGYYVNNDYDD EQLREEPPTKVLIDKVQRNILTDK PRVTKFPINFHPENEQTLGDGPA PTEPFADSVVNGEAPVFLEQPQK LQEIEQFDDSDVNGEAIALLDQPQ NLQET)	10.1111/pce.12299	https://www.uniprot.org/uniprot/Q9LS09
Name				
bHLH122	At1g51140	1-300 (MESEFQQHHFLLHDHQHQRPRN SGLIRYQSAPSSYFSSFGESIEEF LDRPTSPETERILSGFLQTTDTSD NVDSFLHHTFNSDGTEKKPPEVK TEDEDAEIPVTATATAMEVVVSG DGEISVNPEVSIGYVASVSRNKRP REKDDRTPVNNLARHNSSPAGLF SSIDVETAYAAVMKSMGGFGGSN VMSTSNTEASSLTPRSKLLPPTSR AMSPISEVDVKPGFSSRLPPRTLS GGFNRSFGNEGSASSKLTALART QSGGLDQYKTKDEDSASRRPPLA HHMSLPKSLSDIEQLLSDSI)	DAP-seq	https://www.uniprot.org/uniprot/Q9C690
LMI1	**AT5G03790**	140-234 (AKQLEQLYDSLRQEYDVVSREK QMLHDEVKKLRALLRDQGLIKKQI SAGTIKVSGEEDTVEISSVVVAHP RTENMNANQITGGNQVYGQYNN PMLVASSGWPSYP)	DAP-seq	https://www.uniprot.org/uniprot/Q9LZR0

SDG31	**AT3G04380**	196-462 (CCANCKGNCLSADFPCTCARETSGEYAYTKEGLLKEKFLDTCLKMKKEPDSFPKVYCKDCPLERDHDKGTYGKCDGHLIRKFIKECWRKCGCDMQCGNRVVQRGIRCQLQVYFTQEGKGWGLRTLQDLPKGTFICEYIGEILTNTELYDRNVRSSSERHTYPVTLDADWGSEKDLKDEEALCLDATICGNVARFINHRCEDANMIDIPIEIETPDRHYYHIAFFTLRDVKAMDELTWDYMIDFNDKSHPVKAFRCCCGSESCRDRKI)	DAP-seq	https://www.uniprot.org/uniprot/Q8W595
VRN1	At3g18990	100-240 (EINYHSTGLMDSAHNHFKRARLFEDLEDEDAEVIFPSSVYPSPLPESTVPANKGYASSAIQTLFTGPVKAEEPTPTPKIPKKRGRKKKNADPEEINSSAPRDDDPENRSKFYESASARKRTVTAEERERAINAAKTFEPT)	DAP-seq	https://www.uniprot.org/uniprot/Q8L3W1
ATHB13	**AT1G69780**	151-294 (RQFDTLKAENDLLQTHNQKLQAEIMGLKNREQTESINLNKETEGSCSNRSDNSSDNLRLDISTAPPSNDSTLTGGHPPPPQTVGRHFFPPSPATATTTTTTMQFFQNSSSGQSMVKEENSISNMFCAMDDHSGFWPWLDQQQYN)	DAP-seq	https://www.uniprot.org/uniprot/Q8LC03
ABI5	**AT2G36270**	1-340 (MVTRETKLTSEREVESSMAQARHNGGGGGENHPFTSLGRQSSIYSLTLDEFQHALCENGKNFGSMNMDEFLVSIWNAEENNNNQQQAAAAAGSHSVPANHNGFNNNNNNGGEGGVGVFSGGSRGNEDANNKRGIANESSLPRQGSLTLPAPLCRKTVDEVWSEIHRGGGSGNGGDSNGRSSSSNGQNNAQNGGETAARQPTFGEMTLEDFLVKAGVVREHPTNPKPNPNPNQNQNPSSVIPAAAQQQLYGVFQGTGDPSFPGQAMGVGDPSGYAKRTGGGGYQQAPPVQAGVCYGGGVGFGAGGQQMGMVGPLSPVSSDGLGHGQVDNIGGQYGVDMGGL)	DAP-seq	https://www.uniprot.org/uniprot/Q9SJN0
LBD2	**AT1G06280**	(FPARKTKEFQAVHKVFGVSNVQKMVRTVREEDRTKLSDSLTWEALWRQKDPVLGSYGEYRRICEELKLYKSLVHNQPLIGWDNNQRVFNNNSNNKNGLAMTNSSGSGGFSVNNNGVGVNREIVNGGYASRNVQGGWENLKHDQRQQCYAVINNGFKQHYLPL)	DAP-seq	https://www.uniprot.org/uniprot/Q9LNB9
BBX21	**AT1G75540**	110-331 (SSVYKPTSKSSSSSSSNQDFSVPGSSISNPPPLKKPLSAPPQSNKIQPFSKINGGDASVNQWGSTSTISEYLMDTLPGWHVEDFLDSSLPTYGFSKSGDDDGVLPYMEPEDDNNTKRNNNNNNNNNNNTVSLPSKNLGIWVPQIPQTLPSSYPNQYFSQDNNIQFGMYNKETSPEVVSFAPIQN	DAP-seq	https://www.uniprot.org/uniprot/Q9LQZ7

		MKQQGQNNKRWYDDGGFTVPQITPPPLSSNKKFRSFW)		
Other ADs				
TBP1	Plant	MTDQGLEGSNPVDLSKHPSGIVPTLQNIVSTVNLDCKLDLKAIALQARNAEYNPKRFAAVIMRIREPKTTALIFASGKMVCTGAKSEDFSKMAARKYARIVQKLGFPAKFKDFKIQNIVGSCDVKFPIRLEGLAYSHAAFSSYEPELFPGLIYRMKVPKIVLLIFVSGKIVITGAKMRDETYKAFENIYPVLSEFRKIQQ	DAP-Seq	https://www.uniprot.org/uniprot/P28147
BRE1	Plant	700-876 (DIQQGSAYASFLSKKSSRIEDQLRFCTDQFQKLAEDKYQKSVSLENLQKKRADIGNGLEQARSRLEESHSKVEQSRLDYGALELELEIERFNRRRIEEEMEIAKKKVSRLRSLIEGSSAIQKLRQELSEFKEILKCKACNDRPKEVVITKCYHLFCNPCVQKLTGTRQKKCPTCSASFGPNDIKPIYI)	DAP-Seq	https://www.uniprot.org/uniprot/Q8RXD6

Supplementary Table S2. gRNA constructs

Name	Link
AtU6 LacO sgRNA	https://benchling.com/s/seq-tOShh49fIMXhHFR2n00n?m=slm-VN5W2LGHQAlKLQ9ZvgQH
AtWUS 4x gRNA 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 gRNA	https://benchling.com/s/seq-5ziJjbPSXUO7pftspKNn?m=slm-yPNQBDddE3vTFPT7e5wA
FT BlockC gRNA	https://benchling.com/s/seq-PjFkpaHYbWxWtlo9kHf7?m=slm-dhtgd1CTfa1Clvdpywxk
FT BlockE gRNA	https://benchling.com/s/seq-00MeZfV66n40zjyVIC9L?m=slm-vOxkbC9Faka3ax0Zt1Ts
FTgA gRNA	https://benchling.com/s/seq-ZniKsGa2bnk8ZBummIJX?m=slm-bwtcu2gQfWNGLkHDabep
FTgB gRNA	https://benchling.com/s/seq-qPKXU5P5KsZ4M9ekzaPX?m=slm-sjmjHtZt7ZGI5AqzW7qw
FT IntB gRNA	https://benchling.com/s/seq-Lz9NPOLRl0r5fhwCDMUw?m=slm-CWTHhnhIBfr4TKMhU7hJ
FT IntC gRNA	https://benchling.com/s/seq-5CjumRVX0cewolyPVvhQ?m=slm-esyIgOR1PRJkSNgaqKtk
FT IntE gRNA	https://benchling.com/s/seq-UkhXeIHENHQ9mqmUaQLL?m=slm-sTSgfk7BIj0FNqndoEpg
Pap1 gRNA	https://benchling.com/s/seq-gDuRWc02rOPJXxh6z9n8?m=slm-LlXCm27EI7u6OFjqeXKH
WUS gRNA	https://benchling.com/s/seq-wk3d0aJUfeAMj9Hhfv5J?m=slm-Z1fhErOAFrgFMZzfX0zR

Supplementary Table S3. Coding sequences

Name	Link
AtUbi10-dCas9-24xGCN4	https://benchling.com/s/seq-V82BZGxKbmZPsEALuIjz?m=slm-1VUiIlysQL7H07sPVKiB
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

Supplementary Table S4. T-DNA backbones

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

Supplementary Table S5. RT-qPCR primers

Name	Sequence (5' to 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

2.12 Figures

Figure 1 - Screening a library of plant-derived activation domains with the dual luciferase protoplast assay

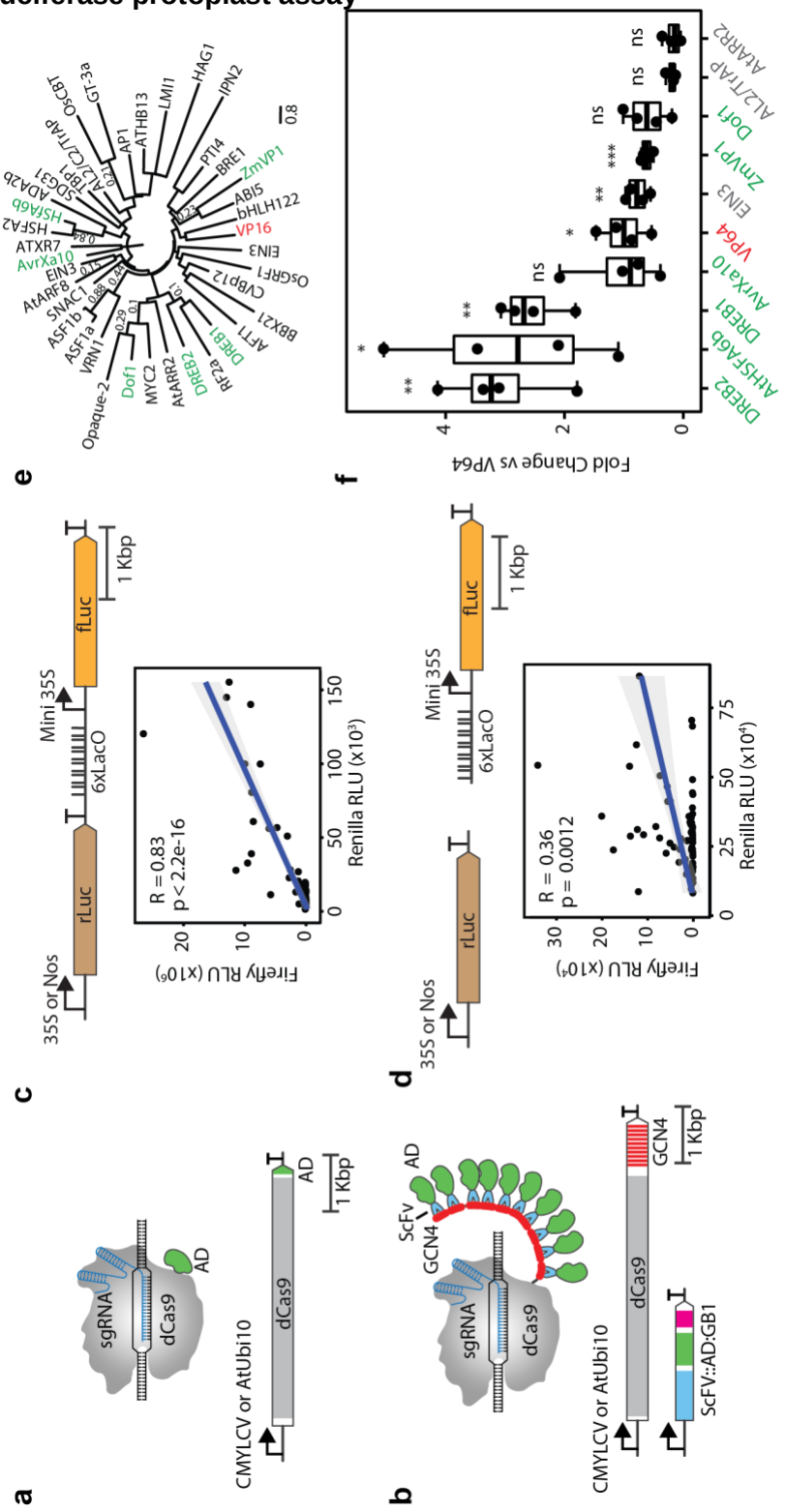


Figure 2. Endogenous gene activation in *Arabidopsis thaliana* protoplasts.

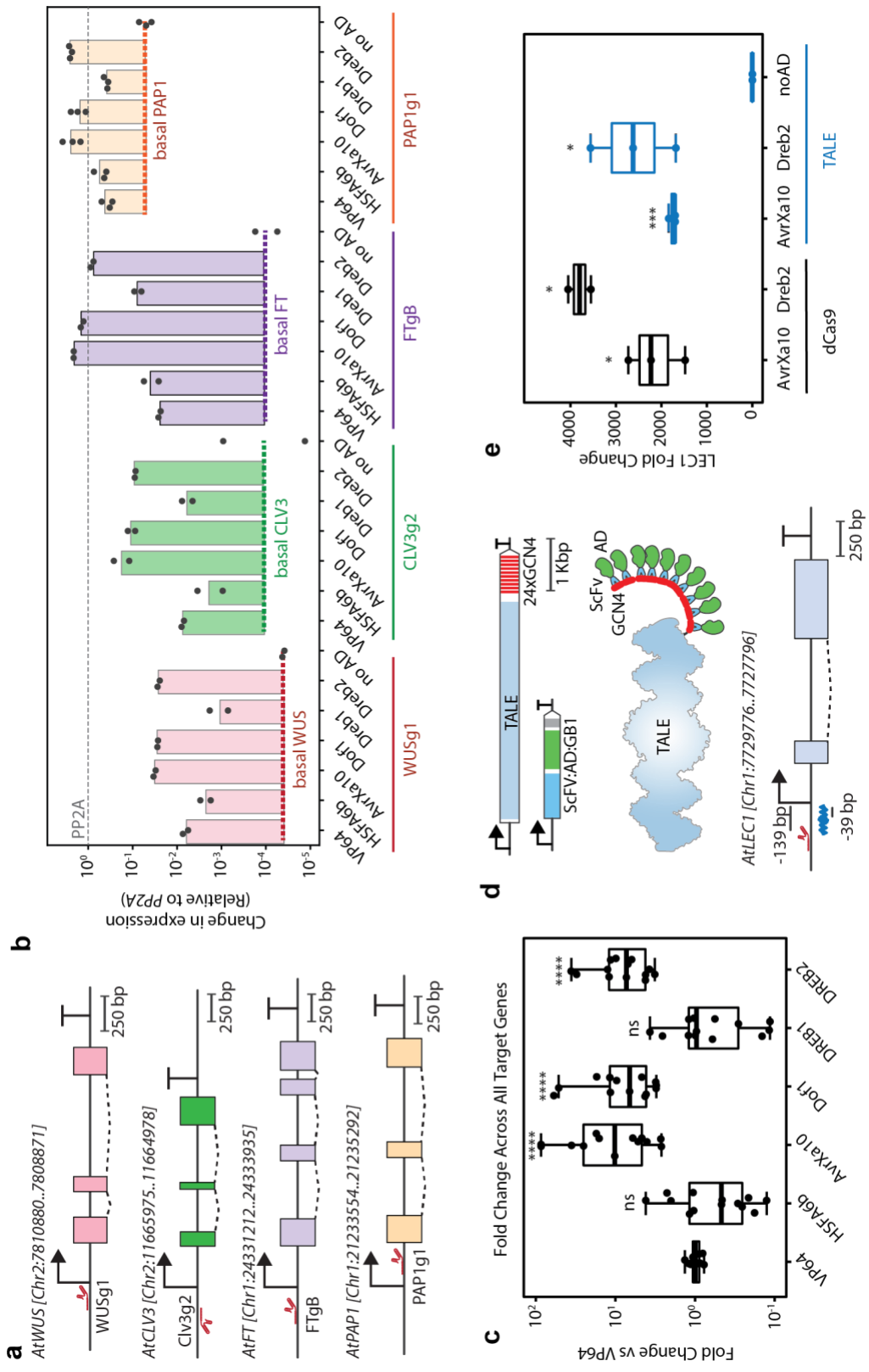


Figure 3. FT gene activation in transgenic plants.

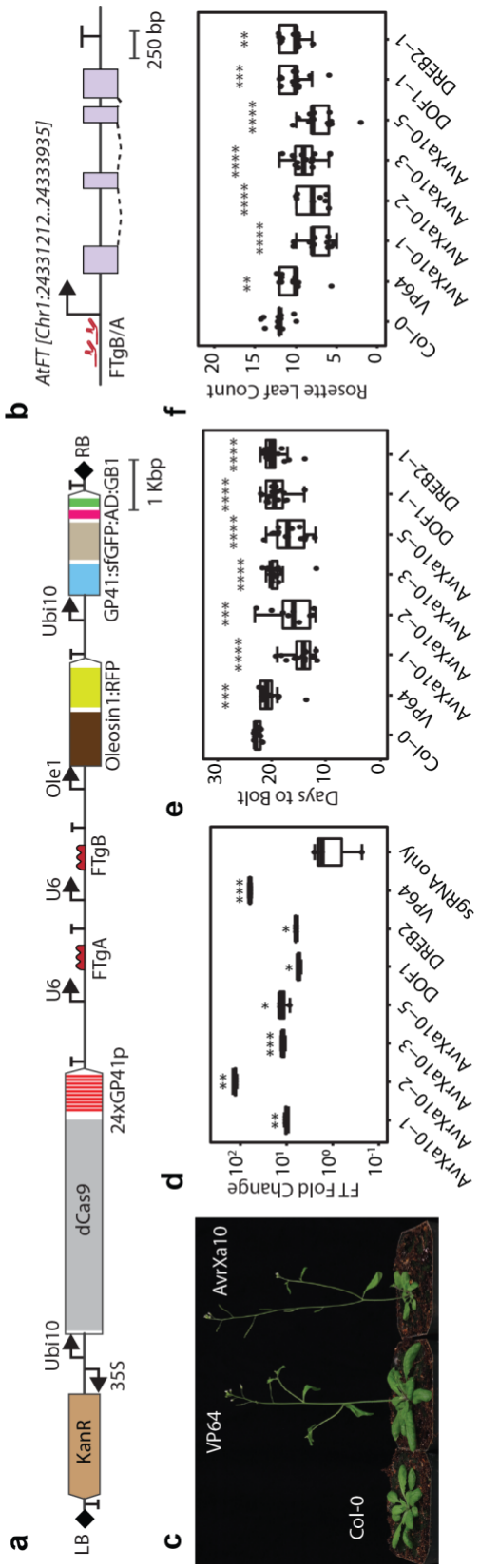
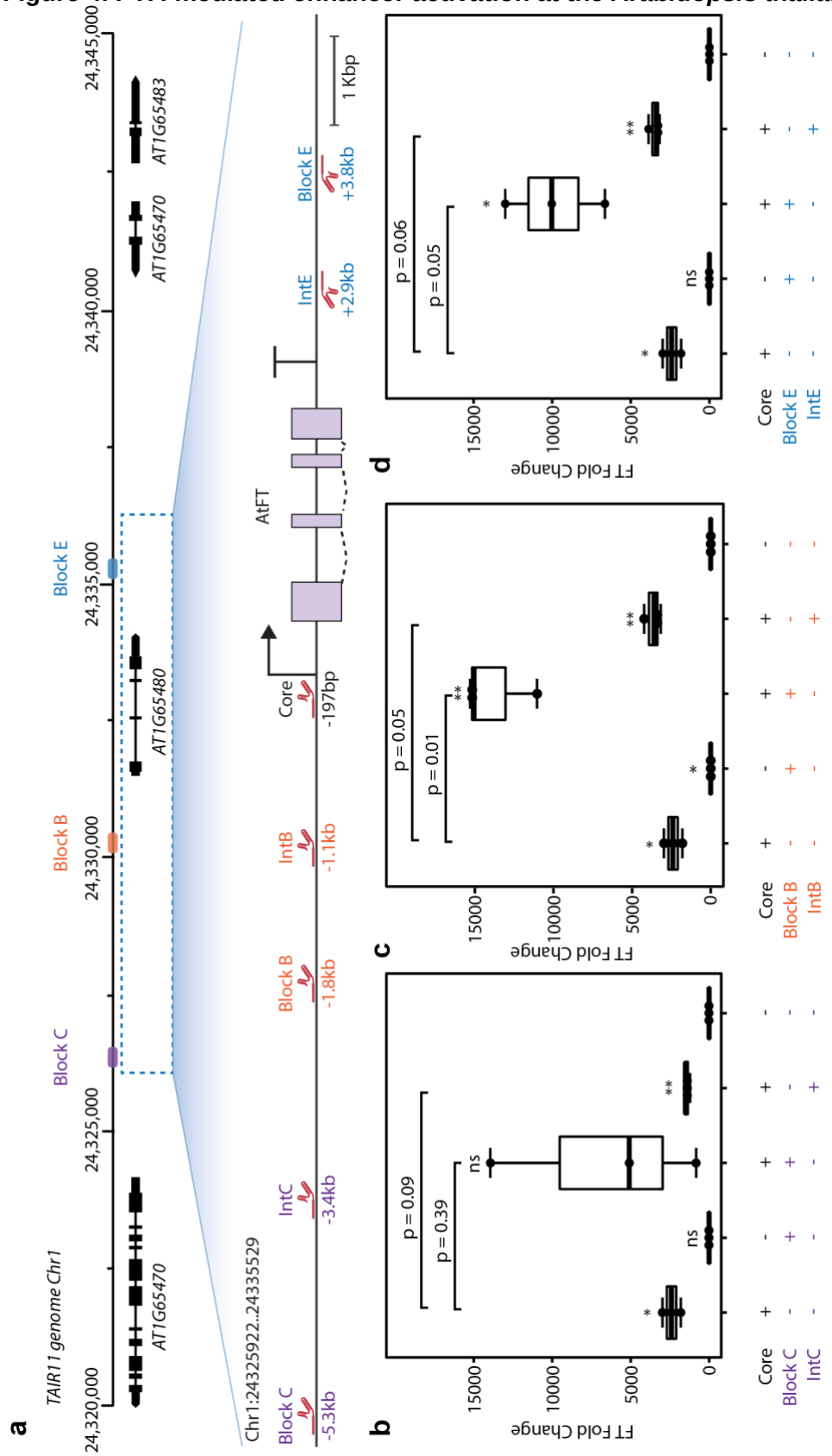
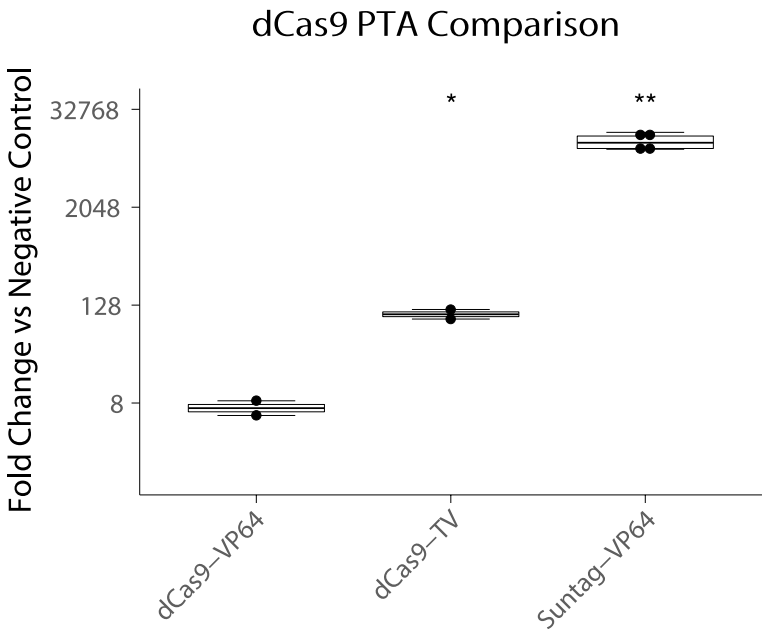


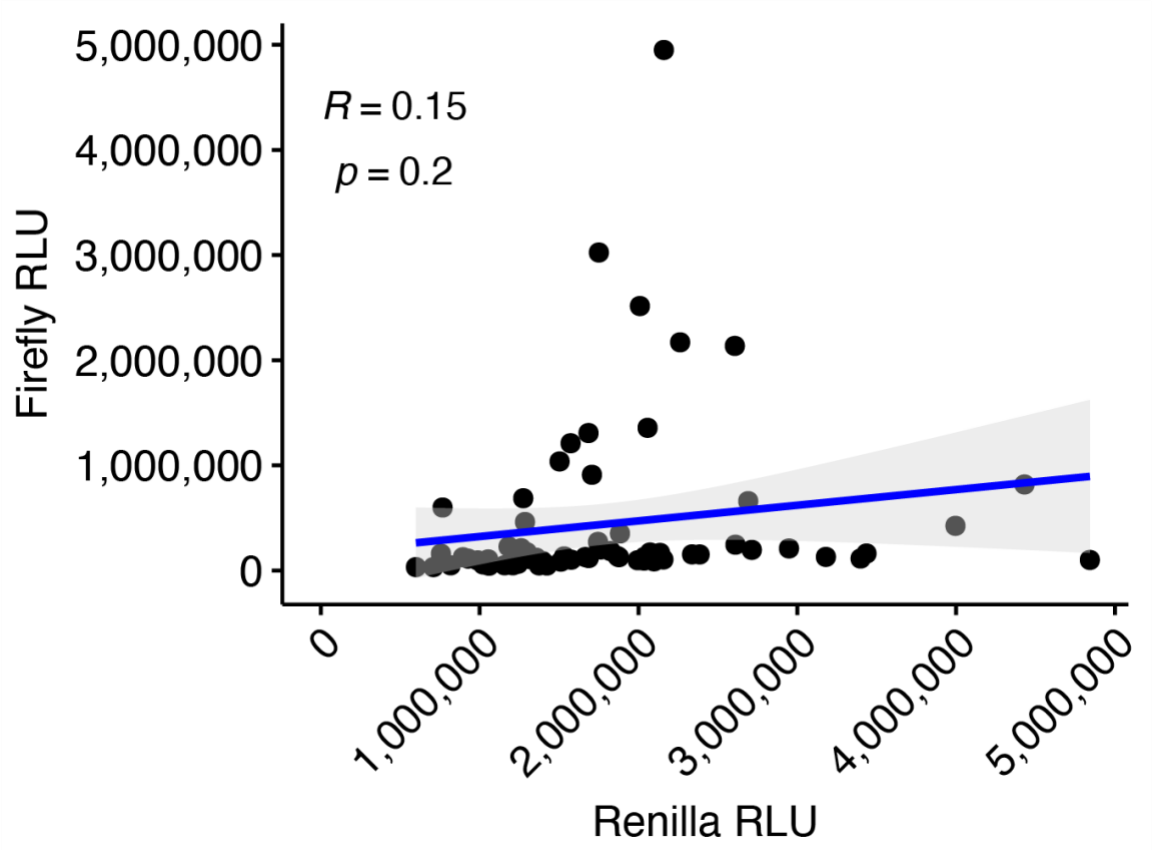
Figure 4. PTA-mediated enhancer activation at the *Arabidopsis thaliana* FT locus.



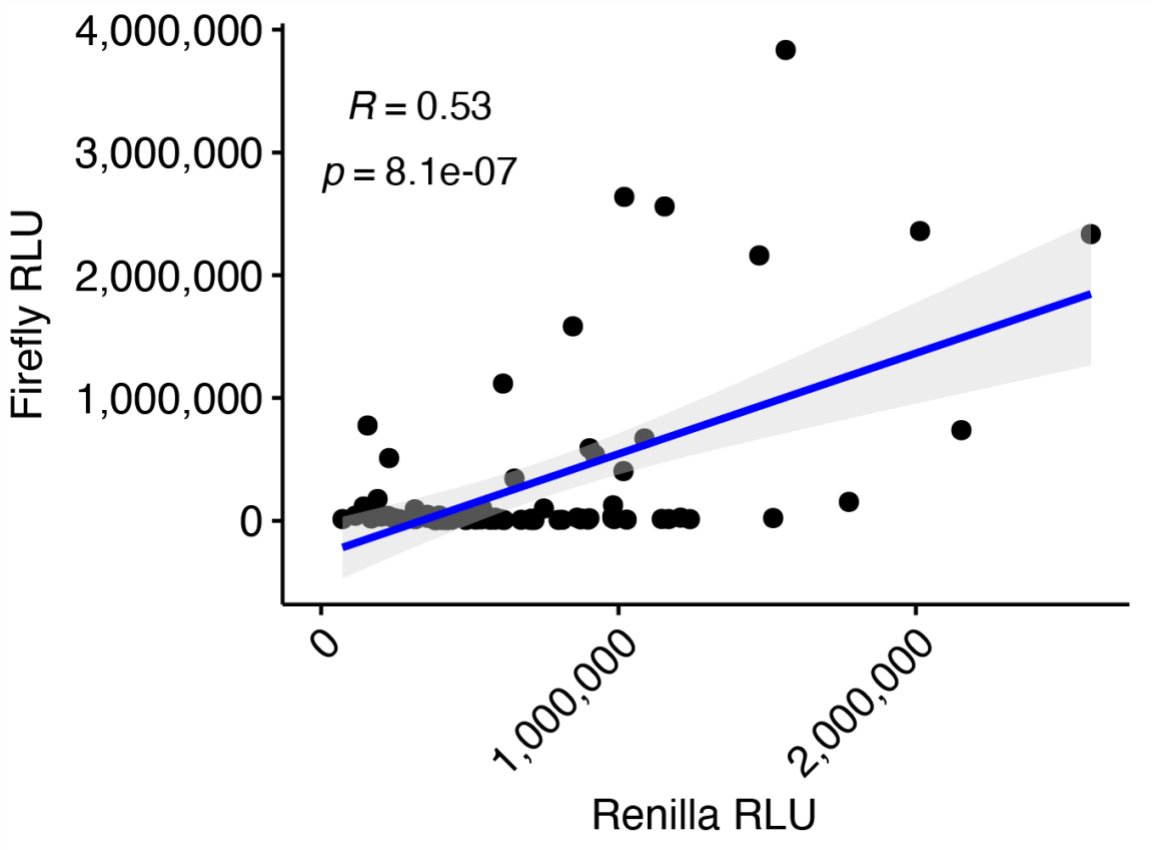
Supplemental Figure 1. Direct fusion PTA vs SunTag PTA comparison in *Arabidopsis thaliana* protoplasts.



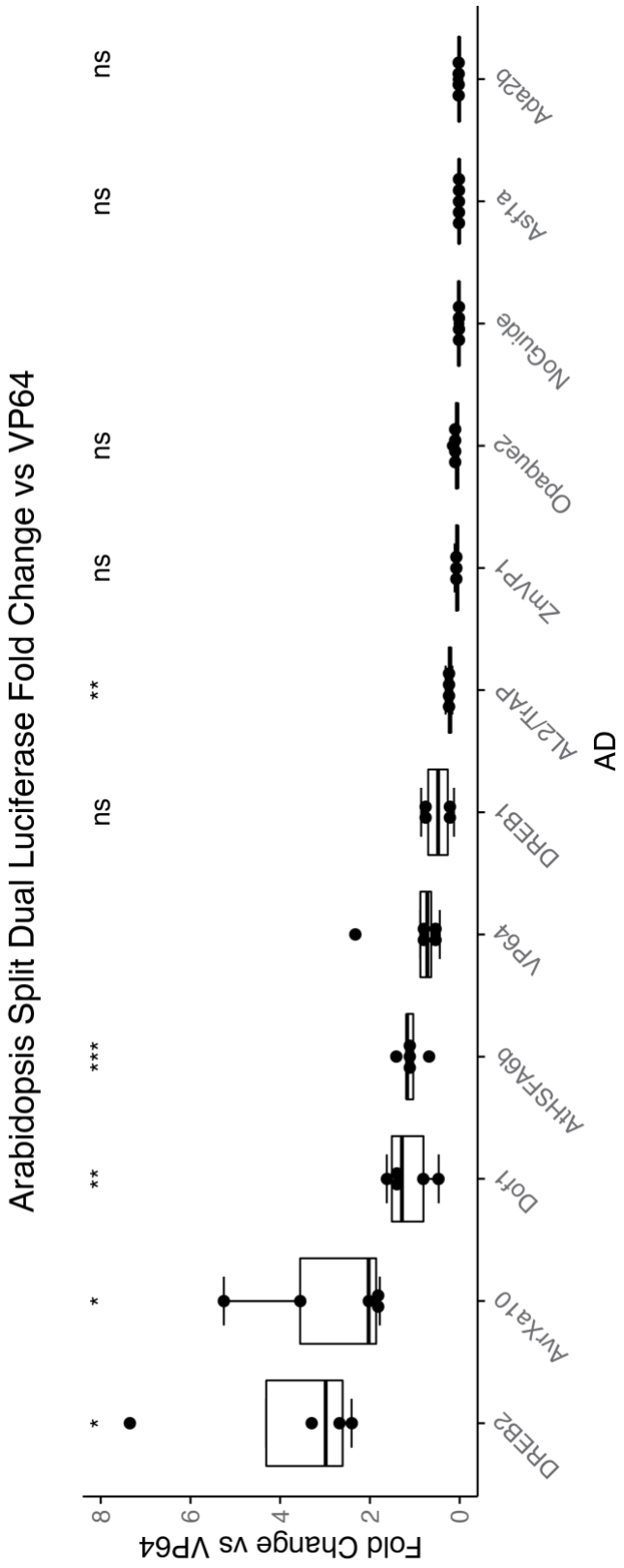
Supplemental Figure 2. 35S Batch One *Setaria viridis* Dual Luciferase Assay.



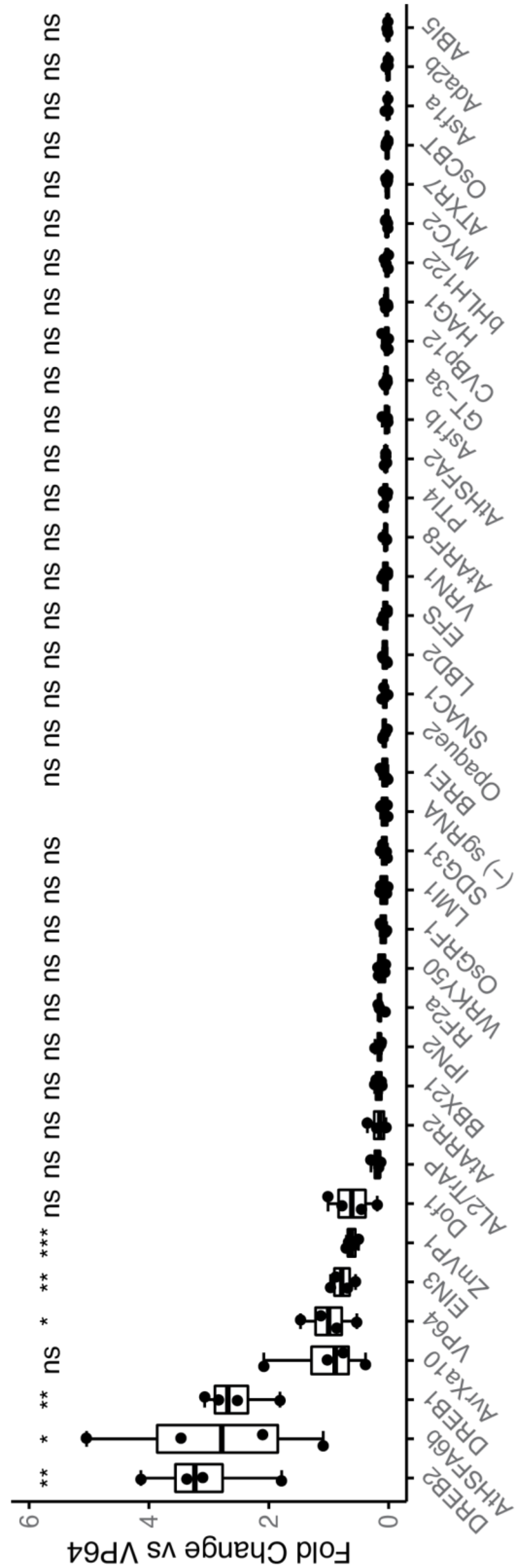
Supplemental Figure 3. 35S Batch Two *Setaria viridis* Dual Luciferase Assay.



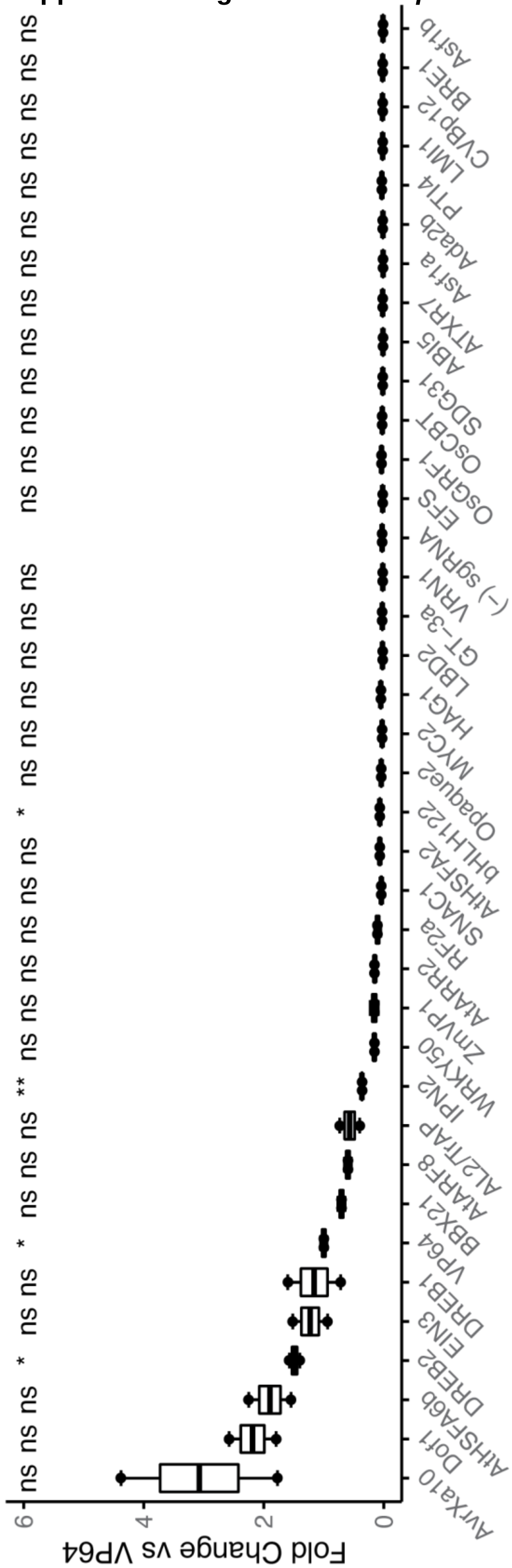
Supplemental Figure 4. *Arabidopsis thaliana* Split Dual Luciferase Assay.



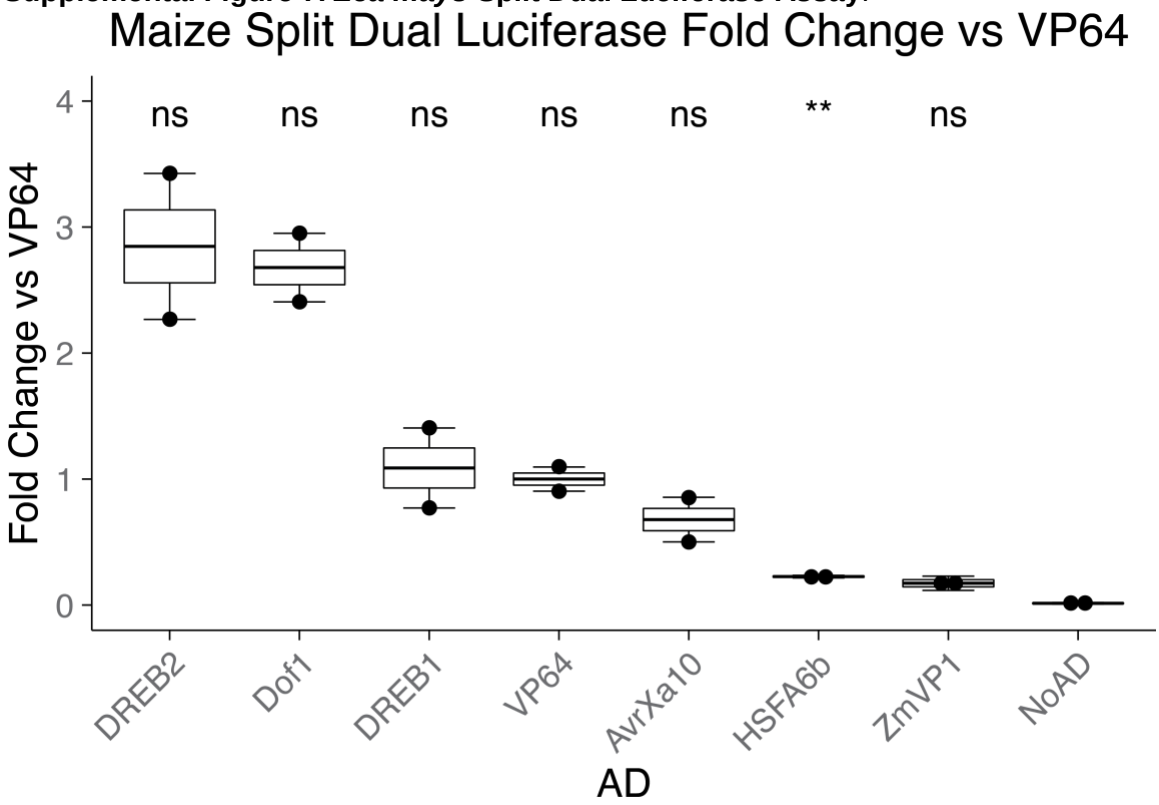
Supplemental Figure 5. *Setaria viridis* Single Dual Luciferase Assay.



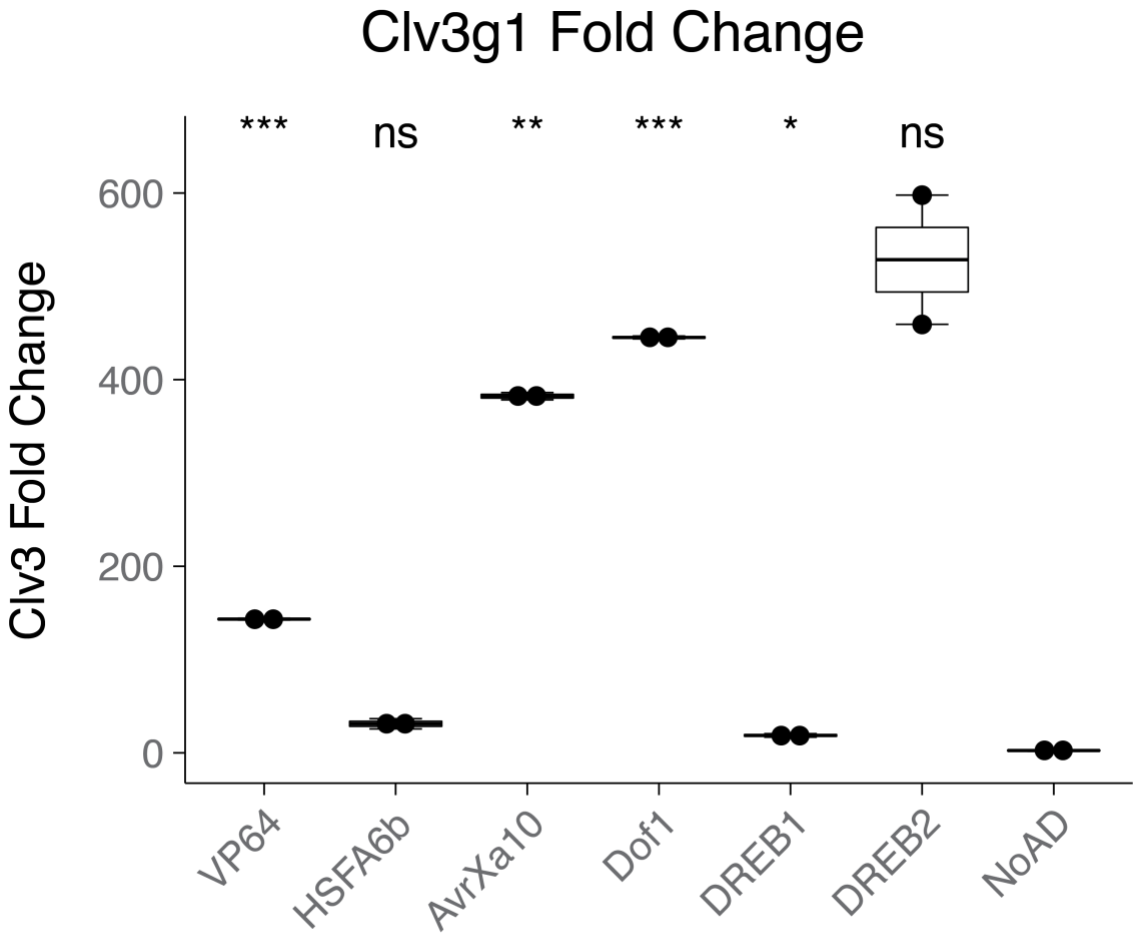
Supplemental Figure 6. ***Arabidopsis thaliana*** **Single Dual Luciferase Assay**



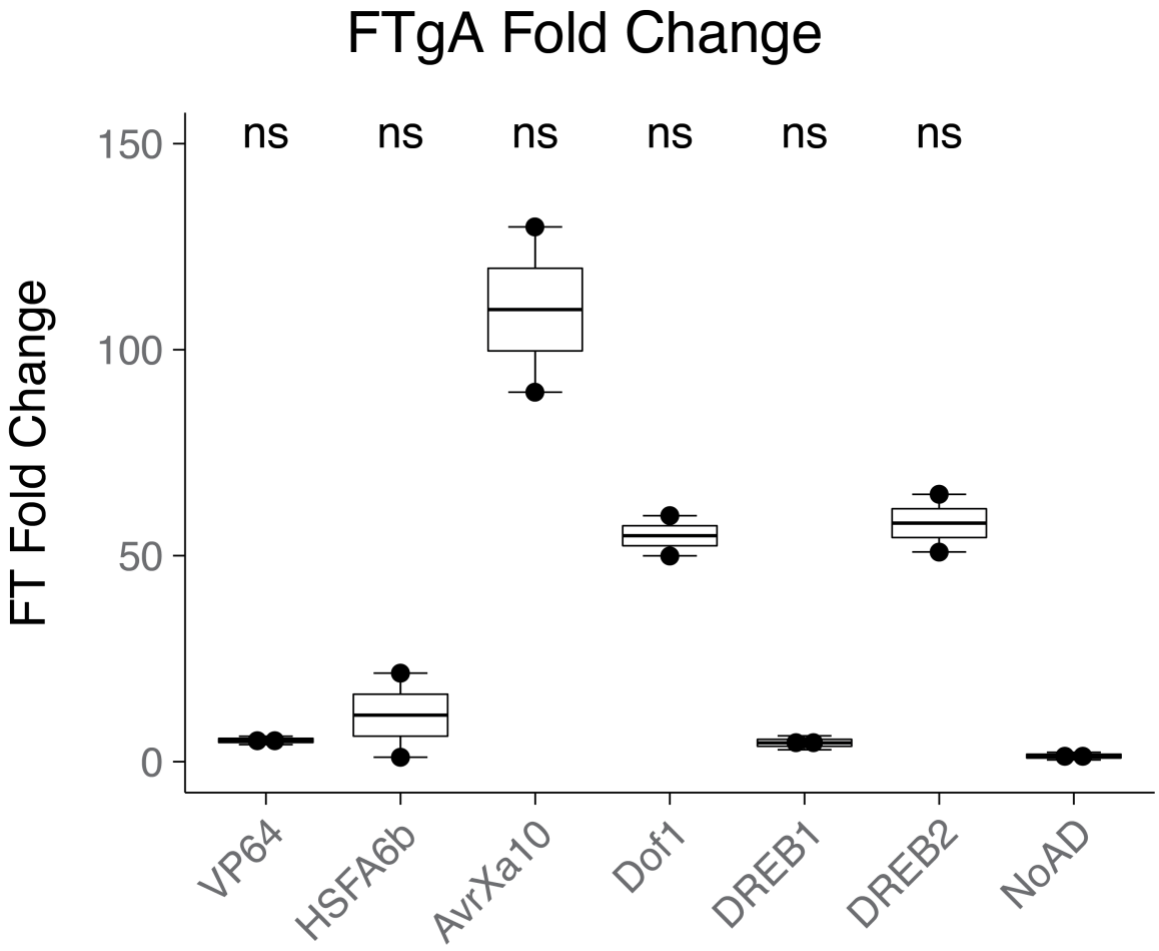
Supplemental Figure 7. *Zea mays* Split Dual Luciferase Assay.



Supplemental Figure 8. *Arabidopsis thaliana* protoplast Clv3 gene activation with gRNA1.



Supplemental Figure 9. *Arabidopsis thaliana* protoplast FT gene activation with FTgA.



Supplemental Figure 10. T1 *Arabidopsis thaliana* Parental Lines

Chapter 3

Building and Implementing Engineered Genetic Incompatibility for Biocontainment in *Arabidopsis thaliana*

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

[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

3.1 – Introduction

Yield loss occurs every year across the world due to a several factors including biotic and abiotic stresses. Yield loss due to biotic stresses such as weeds, animals, or fungi represents a significant contribution to total yield loss[113]. It has been estimated that weeds represent the highest potential yield loss each year at around 34%. In China it has been reported that 35 million hectares of farmland are infested with weeds, reducing crop yields by 12.3 to 16.5%[114].

Typically, weeds can be managed by either mechanical or chemical means. Mechanical weed management centers around soil tillage. For small scale farming this can be accomplished with hand hoeing, while larger scale operations typically require the use of machinery. Mechanical weed removal is effective given the right weather and soil conditions, however as production scale increases so does the cost and time required for effective mechanical weed management.

Thus, over the recent years many farmers have opted for a combination of both mechanical and chemical weed management strategies. Chemical weed management revolves around the use of herbicides to selectively kill weeds in the field. This reduces the amount of labor required to remove weeds as herbicides can be sprayed across a field as opposed to the manual labor required for mechanical removal of weeds from the same field. It is estimated that approximately 50% of the main crops in China undergo herbicide application including rice, wheat, maize, soybeans, and cotton[114]. This same trend holds true in the United States, where farmers deployed glyphosate to 95% of soybean fields and 35% of corn fields by 2006[115].

A key factor to the rapid rise of chemical weed management was the development of genetically modified crops with resistance to herbicides. Monsanto developed the first herbicide resistant crop using a transgenic approach. Glyphosate resistant soybeans were developed by integrating a single gene into the genome of

crops. This gene was called *cp4 epsps*, which encodes for *glyphosate-resistant-5-enolpyruvylshikimate-3-phosphate synthase*[116]. Planting transgenic *cp4 epsps* plants in the field allowed for farmers to simply spray the entire field with glyphosate, resulting in selective death of glyphosate sensitive weeds while the tolerant crops survive.

In comparison to mechanical weed management, these chemical weed management strategies save farmers time and money. However, it also resulted in the increase of herbicide resistant weed populations following widespread adoption. There have been a reported 38 weed species with resistance to glyphosate since its introduction, spanning 37 countries[117]. Of pertinent interest are herbicide resistant weeds that compete with herbicide resistant crops. In 1996, glyphosate resistant soybeans were introduced in the United States. Since then, glyphosate resistant corn, cotton, canola, alfalfa, and sugar beet have been developed and introduced across the United States. Some of the most problematic glyphosate resistant weeds compete with these crops, such as horseweed, and have acquired resistance to a variety of other herbicides in addition to glyphosate resistance[118].

Due to the rapid rise of herbicide resistant weedy populations, there has been skepticism to deploy herbicide resistance in certain crop populations. Of particular concern are crops with known weedy species that can form hybrids in nature. One example of this is *Sorghum bicolor* and the weedy species *Sorghum halepense*. *S. bicolor* is a monocot crop closely related to corn and is of particular interest to the biofuels industry along with being used as feed for livestock. *S. bicolor* and *S. halepense* can form viable hybrid offspring through sexual reproduction in the field[119]. If transformed, engineered *S. bicolor* that hybridizes with *S. halepense* could thus spread the *epsps* transgene from crop to weedy species.

Genetic biocontainment strategies have been developed over the years to prevent this unwanted gene flow from one species to another. Containment is defined as

stopping gene flow form the crop to an unwanted population[120]. Physical containment is the simplest form, in which plants are physically kept apart by a barrier. If carefully managed this can be effective, however, containment of transgenes is difficult to strictly enforce with physical containment given the scale of modern agriculture. Other techniques for containment include cleistogamy, maternal plastid inheritance, and Genetic Use Restriction Technology (GURT), among others.

Cleistogamy is the phenomena of pollination occurring within a closed flower. This can be an effective strategy for containment however even low rates of flower opening can give rise to opportunistic cross-pollination[121]. In addition, if a crop is not capable of self-fertilization this strategy cannot be implemented. Maternal plasmid inheritance is another promising strategy for containment in which transgenes are integrated into the plastid genome. Thus, the transgene cannot be transferred via pollen and only through maternal plasmid inheritance. Unfortunately efficient plastid transformation protocols are limited to a few species, and proteins expressed in the plastids do not undergo the same post-translational modifications[122]. GURT is another biocontainment strategy in which a genetic circuit is designed to result in programmed death when exposed to an inducer. Several iterations of GURT exist, but most relevant to biocontainment is the Recoverable Block of Function (RBF). RBF is comprised of a barnase linked to the gene of interest along with a barstar expressed by an inducible promoter. The barnase will be actively expressed resulting in lethality, until the inducible promoter is active resulting in expression of the barstar to block lethality by the barnase.[123] In this system, the inducible promoter can be a heat shock promoter that is only active under high temperatures (40C) and will not occur naturally. This is a valid biocontainment strategy however there has been significant public pushback to GURT technology over the years.[124]

Next generation genetic biocontainment controls should consider what has been learned over the years to produce robustly effective and trusted biocontainment strategy. One of these next generation containment strategies is Engineered Genetic Incompatibility (EGI).[125] EGI involves the use of Programmable Transcription Activator (PTA) technology to engineer a genetic incompatibility between two subpopulations. In other words, engineering a biological speciation event between two otherwise sexually compatible populations. To do this, genetic relatedness is leveraged to engineer a lethal hybrid mating event only when a weedy species reproduces with a crop species. To do this PTAs are targeted to an endogenous gene promoter driving overexpression of the target gene (Figure 1a, top). When overexpressed this gene causes lethality, for example a developmental gene. The PTA binding site in the promoter is then edited to abolish DNA binding of the PTA (Figure 1a, bottom). This comprises the EGI crop, in which a PTA is expressed monitoring the genome for the unedited DNA binding site (Figure 1b, top). This site remains unedited in the weedy populations, such that when a hybrid mating event occurs the heterozygous offspring will express one copy of the PTA and one copy of the PTA binding site resulting in lethal overexpression and hybrid lethality (Figure 1b, bottom). This concept has been demonstrated in both yeast and Drosophila melanogaster, providing a basis for the method in eukaryotic species.[87,94,126] In this chapter, efforts to implement EGI in the model system Arabidopsis thaliana are described in detail.

3.2 – EGI target gene selection

Central to implementing EGI is selection of an endogenous target gene capable of driving lethal overexpression. Initial work in the model system *Saccharomyces cerevisiae* targeted actin for lethal overexpression. Yeast cells would swell in size and eventually lyse due to actin overexpression. EGI in *Drosophila melanogaster* used the target gene *pyramis* for lethal overexpression. The *pyramis* gene is a fibroblast growth

factor ligand for the receptor *heartless*. PTA-mediated ectopic expression of *pyramis* results in defects of both spreading and mesoderm differentiation.[127] Thus there is precedent for different classes of EGI target genes with lethality occurring through multiple mechanisms.

When searching for EGI target genes we leaned heavily on the literature. Previous studies had implicated the key developmental genes *wuschel (wus)* and *leafy cotyledon 1 (lec1)* as having lethal overexpression phenotypes.[128,129] These genes are involved in early embryo formation and development. Ectopic postembryonic expression of these gene products results in aberrant somatic embryogenesis and the formation of embryo-like structures.[130] This eventually leads to lethality, as proper development and differentiation of somatic tissues cannot occur. There are many potential EGI candidate genes involved in early plant development that could trigger ectopic embryogenesis upon PTA-mediated activation, such as *shoot merisitemless (stm).*[131] One could continue adding EGI candidate genes from these pathways, however, we stopped with *wus*, *lec1*, and *stm* as three developmental EGI target genes to test in plants.

In addition, we explored other mechanisms of lethality outside of plant development. One alternative pathway is programmed cell death. In plants, leaf senescence is a form of programmed cell death in which molecular cues produced by stress events such as wounding or infection initiate a programmed cell death response. In *Arabidopsis thaliana* it has been demonstrated that the *UBA2a-c* genes are activated upon mechanical wounding of leaf tissue.[132] The *UBA2* family of genes are nuclear ribonucleoprotein RNA binding proteins that play a critical role in post-transcriptional processing of RNA. The *UBA2a-c* genes were added to the EGI candidate gene list because of observed lethality in transgenic Arabidopsis plants carrying a 35S-*Uba2* cassette for constitutive expression.[132]

Other EGI target genes were identified based on the perturbation of key cellular processes for observed lethality. Among these genes are *myb21* and *rsz33*. *Myb21* was first identified as a gene regulated by *cop1*, a key regulator of light-response signaling pathways. *Cop1* encodes a protein containing a RING finger motif, a coiled-coil motif, and WD40 repeats.[133] *Cop1* nuclear accumulation is regulated by the Cop9 signalosome, which bears resemblance to the 26S proteasome.[134] Mutations in the Cop genes tend to result in disrupted photomorphogenic development.[135] *Myb21* is a downstream target of *Cop1* and is a transcription factor responsible for the early stages of photomorphogenesis. Ectopic expression of *Myb21* results in dwarfism, flower malformation, and seedling lethality.[136]

Splicing is another key process that involves the removal of introns and ligation of exons by the spliceosome to form a mature RNA transcript. While there are some sequence motifs that identify intron-exon boundaries, there is flexibility in where the spliceosome will cut and ligate a transcript. This variance in splicing is termed alternative splicing, which leads to the formation of alternative transcript isoforms. The spliceosome itself is a large complex of RNA and proteins, with one class of proteins in the complex being the arg/ser-rich (SR) proteins.[137] SR proteins are conserved throughout metazoans and play a central role in the splicing process.[138] SR proteins contain both RNA binding domains and protein-protein interaction motifs, making SR proteins essential for splice site selection and spliceosome assembly. In addition, tissue specific expression of different SR proteins can result in alternative splicing of transcript isoforms.[139] While studies have shown that SR proteins can be deleted due to their functional redundancy, ectopic expression of SR proteins can interfere with proper splicing and perturb development.[140] In *A. thaliana* ectopic expression of the gene *rsz33* results in pleiotropic phenotypic changes, most notably in early development. Twin embryo formation, perturbed cotyledon growth producing a single cotyledon, and

abnormal division in the suspensor at globular stage were all noted upon ectopic *rsz33* expression.[141]

Taken together, we identified a set of EGI target genes spanning a diverse array of cellular processes necessary for survival and reproduction in plant systems. We first set out to test these target genes in the model system *A. thaliana* for observed lethality. Upon EGI demonstration in the model system A. thaliana, a similar conserved gene or gene family can be targeted in crop species.

3.3 – Targeting pro-apoptotic leaf senescence genes for lethal overexpression

We first set out to activate the genes *uba2b* for lethal overexpression in the model system Arabidopsis thaliana. To do so, we built T-DNA vectors to express an engineered Transcription Activator Like Effector (TALE) DNA binding domain fused to an activation domain (AD). This TALE-AD fusion was engineered to bind the promoter region of the *uba2b* target gene, specifically within 500bp of the transcription start site (TSS).

The TALE DNA binding domain can be engineered based on the TALE code.[142–145] A single TALE repeat is a stretch of 34 amino acids capable of binding a single nucleotide.[143] Sets of 20 TALE repeats can be cloned into a TALE backbone, resulting in a engineered TALE capable of binding a 20 nucleotide genomic target site. Nucleotide specificity of a single TALE repeat is conferred by the Repeat Variable Disresidue (RVD), two amino acids at the 12th and 13th positions of the TALE repeat that recognize a specific nucleotide.[143] The amino acids 'NI' prefers to bind the 'A' nucleotide, 'HD' prefers to bind the 'C' nucleotide, 'NG' prefers to bind to bind the 'T' nucleotide, and 'NN' prefers to bind the 'G' nucleotide.[146] The only targeting constraint for TALE DNA binding domains is that the 5' nucleotide immediately upstream the target site must be a 'T'.

Given these constraints, TALE DNA binding domains were generated to target the *uba2b* promoter region (AT2G41060). Many of these domains targeted a 100 bp

region starting 40bp upstream of the TSS (Figure 2a). TDNA vectors were built containing the TALE-VP64 or TALE-VPR PTAs (Figure 2b). Wild type Col-0 *A. thaliana* plants were floral dipped with Agrobacterium tumefaciens carrying *uba2b* TALE TDNAs along with a selectable marker. Transgenic lines were selected using antibiotic resistant media, and phenotypes were analyzed. Across all 14 transgenic lines, phenotypes mirrored wild type plants (Figure 1c). From this result, we concluded that a transient assay was needed to first determine if the engineered PTAs could activate the target gene prior to investing time and resources in transgenic lines.

3.4 – Transient agrobacterium assays to determine PTA guide RNA activity for transgenesis with dCas9-TV*

This section includes excerpts from previously published work, republished with permissions.

Fast-TrACC: A Rapid Method for Delivering and Testing Gene Editing Reagents in Somatic Plant Cells

Ryan A. Nasti[1,2,3], Matthew H. Zinselmeier[1,2,3], Macy Vollbrecht[1,2,3], Michael F. Maher[1,2,3,4], and Daniel F. Voytas[1,2,3]

1 Department of Genetics, Cell Biology, and Development, University of Minnesota, St. Paul, MN, United States 2 Center for Precision Plant Genomics, University of Minnesota, St. Paul, MN, United States 3 Center for Genome Engineering, University of Minnesota, St. Paul, MN, Plant and Microbial Biology Graduate Program, University of Minnesota, St. Paul, MN, United States.

Producing a gene edited plant requires considerable time, often from 6 to 9 months.[147] Over this time period, significant effort must be put forth to identify edited

cells in culture and induce them to form shoots and roots. Because of this investment in time and labor, it is important to know at the onset of an experiment whether the gene editing reagents can effectively create the desired genetic change. Typically, reagents are tested using transient assays to determine reagent efficacy within a shorter timescale. By comparing several different reagents in this manner, the most efficient one can be selected and used to generate the gene edited plant. Currently, the most common transient delivery systems involve protoplasts[148] or leaf infiltrations.[149,150] While both are effective, each has its own associated drawbacks. Protoplast isolation, where one removes the cell wall from plant cells, allows for transient transformation by chemical methods or electroporation. Isolating protoplasts is technically challenging and places the cells in an unnatural environment. On the other hand, leaf infiltration, performed by perfusion of *Agrobacterium tumefaciens* into a leaf with a needless syringe, is simple to perform but works with a limited number of plants, and time is required to grow plants to the proper stage for infiltration.

An alternative method, called AGROBEST, was developed for transient expression of transgenes in *Arabidopsis thaliana*.[151] In this method Agrobacterium cultures are placed in media to promote expression of the *vir* genes, thereby improving the efficiency of T-DNA transfer to plant cells. With this increase in *vir* expression, one can deliver a given T-DNA cargo by simply co-culturing Arabidopsis seedlings with the treated bacterial culture. We sought to use this approach to deliver T-DNA cargo to *Nicotiana benthamiana* seedlings, however, in order to achieve transformation, it was necessary to make changes to the concentration of bacteria used and the length of time the seedlings and bacteria were co-cultured.[152] Specifically, increasing the Agrobacterium concentration and shortening co-culture times resulted in improvements in transgene delivery. This altered method, fast treated Agrobacterium co-culture (Fast-

TrACC), was used to deliver developmental regulators to *N. benthamiana* seedlings to induce *de novo* meristems to create either transgenic or gene edited shoots.[152]

The success of Fast-TrACC in *N. benthamiana* suggested that it might be generally useful as a transient DNA delivery method. Fast-TrACC was used to efficiently deliver transgenes to other related species, including tomato (*Solanum lycopersicum*), potato (*Solanum tuberosum*), pepper (*Capsicum chinense*), and eggplant (*Solanum melongena*). Fast-TrACC was also used to compare the activity of various promoters in these species using a luciferase reporter, and we demonstrate that Fast-TrACC can quickly assess the activity of gene editing reagents at endogenous chromosomal targets. With relative ease, Fast-TrACC makes it possible to identify the reagents with highest activity prior to generating a gene edited plant line.

We leveraged the Fast-TrACC to first screen TDNAs for activity in plant cells prior to generating transgenic lines. At this point, the transition was made from TALE-based PTAs to dCas9-based PTAs. The reason for this was the simplicity and speed of guide RNA design, allowing for rapid vector construction and subsequent testing in plant cells as opposed to the more laborious TALE assembly pipelines.

The first vectors developed were a series of sgRNAs targeting the *wuschel (wus)* promoter for gene activation by dCas9-TV. The dCas9-TV activator was demonstrated to be a strong plant PTA by fusing of six TAL ADs and two VP64 ADs to a single dCas9 (Figure 3a, top)[14]. However, we observed toxicity when growing the dCas9-TV plasmid in *E. coli*. We hypothesized this toxicity could be due to the repetitive sequences of the TV activation domain. To reduce this, we ordered a codon optimized synthetic gene block for TV which we will refer to as TV* (Figure 3a, bottom). This TV* domain was designed using a sliding window to reduce repetitive codon usage, resulting in sequence diversification across the TV* domain. In addition, we cloned a potato IV2 intron into the dCas9 gene to improve stability of the construct. Prokaryotes do not perform splicing,

and insertion of the IV2 intron would result in an out of from dCas9 protein product. We observed better plasmid growth and stability in E.coli following these changes, as indicated by comparing colony PCR across the TV and TV* domains following liquid culture outgrowth (Figure 3b).

This PTA was cloned into a TDNA alongside a guide RNA under the expression of a U6 polymerase III promoter. *Agrobacterium tumefaciens* strain GV3101 was transformed and 1 week old *Arabidopsis thaliana* seedlings were co-cultured according to the Fast-TrACC protocol (Figure 3c). The first promoter targeted was the *Arabidopsis thaliana* gene *Wuschel* (*WUS*). Guide RNAs were designed to target the core promoter region, spanning -50bp to -1000bp upstream of the annotated TSS (Figure 3d). Each sgRNA was cloned into its own TDNA, and seedlings with strong transient expression were selected two days following co-culture for RNA extraction (Figure 3e). We then DNAse treated samples and performed RT-qPCR to quantify *WUS* gene activation using the delta delta Ct method. From these experiments we observed the three sgRNAs located closest to the TSS produced the strongest *WUS* gene activation data (Figure 3f).

Given the results of the transient Fast-TrACC assay, we reasoned that WUS overexpression with dCas9-TV and sgRNAs 4-6 may also occur in transgenic plants. Floral dips were performed as described previously, and T1 transgenic lines were identified by selection in Kanamycin media. Unfortunately, all plants again appeared wild type three weeks post germination (Figure 3f). Given these results it was concluded that high throughput transient expression systems may be necessary to identify effective constructs for EGI.

3.5 – Protoplast transient assays demonstrating SunTag-mediated EGI target gene activation

Shortly after the dCas9-TV* transgenic lines failed to produce a phenotype, the SunTag activation system was demonstrated to be an efficient activator of gene

expression in *Arabidopsis thaliana*[60]. This system utilized the AD scaffolding method, in which multiple ADs can be recruited to a single programmable DNA binding domain. This was accomplished by fusing and multimeric epitope tail to dCas9 and fusing the AD to an scFv antibody. Each epitope is recognized by an antibody, resulting in recruitment of multiple scFv antibodies to a single dCas9 molecule. In the case of Papikian et al, ten GCN4 epitope repeats were fused to dCas9, resulting in up to ten VP64 ADs recruited to a single dCas9.

An in-house version of the SunTag system was developed in which 24 copies of the GCN4 epitope were fused to dCas9. Protoplast isolation and transformation assays were used to compare the activity of dCas9-TV* to the dCas9-SunTag system (Figure 4a, top). Multi-guide expression arrays for an EGI target gene of interest were cloned to obtain maximal gene activation. These multi-guide arrays use a polymerase II promoter to drive expression of an mRNA molecular containing sgRNAs flanked by tRNA sequences (Figure 4a, bottom). The tRNA sequences are self-processing and will liberate the sgRNAs post-transcriptionally, resulting in simultaneous expression of four sgRNAs from a single transcript. Multi guide arrays were built for the EGI target genes WUS, Lec1, and STM. The sgRNA binding sites for each respective core promoter span 20-300bp upstream of the annotated TSS (Figure 4b).

Arabidopsis thaliana protoplasts were isolated from 3-week-old plants and transformed with plasmids encoding either SunTag (dCas9-24xGCN4 and scFv-AD) or dCas9-TV*, along with the multi-guide array. The dCas9-24xGCN4 and dCas9-TV* were driven by the AtUbi10 promoter, while the scFv-AD and multi-guide array were driven by 35S promoters for strong constitutive expression in mesophyll protoplasts. We tested the SunTag system using the conventional VP64 activation domain fused to scFv, and the DREB2 activation domain described in Chapter 2. RNA was isolated 24h after transformation followed by RT-qPCR to detect the target gene of interest. We measured

gene activation for all three EGI target genes WUS, LEC1, and STM (Figure 4c-f). Maximal transcript abundance observed across all three target genes was roughly equivalent to the housekeeping gene PP2A when transforming with the SunTag system, particularly with the DREB2 activation domain (Figure 4C). Fold change values versus a negative control also reflect this trend of maximal activation using the SunTag activation system where statistical significance was reached for all three target genes (Figure 4d-f). Given these results, we moved forward with building transformation vectors for testing EGI target genes for lethal overexpression *in planta*.

3.6 – Demonstration of EGI target gene overexpression *in planta* and building the EGI line

The next objective was to demonstrate activation of EGI target genes in transgenic plants. It is unclear how PTA-mediated activation of EGI target genes will manifest in transgenic lines. In addition, plant transformation with Agrobacterium results in random integration of the TDNA leading to variation in transgene expression. This is typically overcome by placing a herbicide resistance gene on the TDNA allowing for selection of lines strongly expressing the transgenic cassette. A straightforward approach to demonstrate target gene lethality is to include a herbicide resistance gene, PTA, and sgRNAs on a single TDNA, followed by selection for strong expressing lines. However, it may be difficult to differentiate between non-transgenic lines lacking the transgene and strong expressing PTA lines as both may result in lethality.

To avoid this discrepancy, transgenic lines were made to separate the PTA and sgRNA. Specifically, TDNAs were generated to contain either the PTA + herbicide resistance or the sgRNAs + herbicide resistance (Figure 5a). The PTA line was linked to a firefly luciferase and the sgRNA line was linked to a GFP for visible screening of transgenic progeny following selection (Figure 5a). In addition, different resistance genes were used for the PTA line and sgRNA line. This will allow for crossing of the PTA line to

the sgRNA in order to observe EGI target gene overexpression independent of transgenesis. Using selection for both cassettes will be performed to identify successful cross progeny, but also to ensure strong expression of both transgenes in the hybrid offspring following crosses.

TDNA vectors were generated and floral dipped into Arabidopsis thaliana. T1 seeds were collected and plants were subsequently placed on selection to identify transgenic lines. For the SunTag PTA line the dCas9-24xGCN4 was under the control of the AtUbi10 promoter while the scFv-VP64 was driven by the 35S CaMV promoter (Figure 5a). Many SunTag PTA lines were identified following selection on hygromycin. Expression analysis was subsequently performed by extracting RNA and performing RT-qPCR to quantify gene expression of both the dCas9-24xGCN and the scFv-VP64 (Figure 5c). Interestingly, the dCas9 tended to be expressed higher than the scFv across many transgenic lines. Visual confirmation of Luciferase expression was performed in the T2 generation (Figure 5d).

The EGI multi-guide arrays were cloned from the protoplast assays as described previously, in which four sgRNAs target either the *WUS* or *LEC1* core promoter (Figure 5a-b). These sgRNA lines contained a multi-guide array for the EGI target gene driven by the 35S promoter along with a GFP under the control of the Rbcs3b promoter for visible screening. Lines were identified by selection on Kanamycin, and visual confirmation of GFP expression was detected in the T2 generation (Figure 5d).

Crosses were then performed with mature T2 parents. The SunTag line was used as the maternal parent, while the sgRNA was used as the pollen donor. Crosses were performed with the Lec1 multi-guide array and the WUS multi guide array lines crossed to SunTag PTA lines. T3 hybrid seeds were collected from the T2 parents used in crosses, and T3 seedlings were plated on double Kanamycin + Hygromycin selection

media. This allows for confirmation of a successful cross, and strong expression of both transgenes in the hybrid offspring.

Phenotypes were observed in both crosses. For the WUS crosses, we observed many plants arrest during development one week post germination (Figure 5e). Nearly all plants lacked chlorophyll. GFP expression was detectable in some individuals, particularly early in development. As time progressed the GFP signal faded, and many plants died on the double selection media. In some instances, clear sectors of GFP positive cells could be identified on older seedlings approximately three weeks post germination (Figure 5f). These patches of GFP positive cells displayed many characteristics of the expected WUS overexpression phenotype including callous-like growth from leaf tissue in the absence of any hormones (Figure 5f).

For the Lec1 crosses, we observed many plants with abnormal development. Strong GFP expression was detected in roughly half of the T3 seedlings on double selection media one week post germination (Figure 5e). Of those GFP positive seedlings, many of them displayed abnormal root development. The cotyledons and true leaves developed as expected, however, roots failed to develop in nearly all the GFP positive plants approximately three weeks post germination (Figure 5f). Abnormal structures were produced instead of roots which appeared to show some characteristics of roots, but never grew into the media. These abnormal root-like structures also had ubiquitous GFP expression, indicating they may be result of LEC1 overexpression (Figure 5f).

A handful of hybrid progeny were sacrificed one week post germination for expression analysis. For each cross performed, two GFP-positive seedlings and two GFP-negative seedlings were sacrificed for RNA extraction. RT-qPCR was performed on these seedlings to detect expression of four genes in hybrid progeny as compared to a control line expressing only a PTA. Expression of the SunTag-PTA was analyzed first, in

which all lines showed strong expression of the PTA (Figure 5g-h). Next, GFP expression was quantified to confirm expression of the sgRNA. This locus exhibited significant variation across hybrid progeny. In the WUS hybrids, GFP expression was detectable but low in comparison the GFP expression in the LEC1 hybrids (Figure 5g-h). Gene activation of the EGI target genes WUS and LEC1 were detected in hybrids offspring (Figure 5g-h). EGI target gene activation appeared to be correlated with GFP expression, as high expressing GFP lines from the LEC1 hybrid progeny showed the highest EGI target gene activation (Figure 5h).

These results indicated that both LEC1 and WUS showed significant promise as EGI candidate, particularly the *WUS* gene as none of the cross progeny were able to develop past the first set of true leaf development. In order to make the EGI line using *WUS* as the target, promoter mutations are necessary to protect the EGI organism from self-targeting. The multi-guide array used for *WUS* overexpression in the previous experiments expressed four sgRNAs all binding within a 200bp window 50bp upstream of the *WUS* TSS. If these sgRNAs were co-expressed alongside Cas9, there the potential for a large promoter deletion spanning the entire region between *WUS* sgRNA1 and sgRNA 4 (Figure 5, top).[153] This could be problematic as core promoter deletions, particularly those spanning a Transcription Start Site (TSS), can cause significant changes to gene expression and potentially phenocopy a knockout mutation of the gene itself.[154]

It is preferred to use a single sgRNA for mutagenesis, as the likelihood of generating a large core promoter deletion using a single sgRNA is lower than when using multiple sgRNAs.[155,156] Small indel mutations within a promoter region can be tolerated without significantly compromising native gene expression, given that they do not occur in a transcription factor binding site.[157,158] Transient *A. thaliana* protoplast isolation and transformation was performed to identify which of the four WUS sgRNAs

contributed the most to gene activation. It was clear that sgRNA1, located closest to the TSS, produced the strongest gene activation more than three orders of magnitude higher than the other three sgRNAs targeting the same promoter (Figure 5i).

Following this result, TDNA vectors were designed to express a catalytically active version of Cas9 along with WUS sgRNA1 for mutagenesis of the sgRNA binding site. T1 plants were identified using the pFAST system by selecting for RFP positive seedlings carrying the mutagenesis TDNA. Leaf tissue was collected for DNA extraction, followed by PCR of the WUS promoter containing the sgRNA target site. Mutations were readily detected by Sanger sequencing in the T1 generation, with frequencies reaching up to 99% across 30 T1 lines that were screened. We then selected for RFP negative T2 seedlings in the next generation, followed by DNA isolation and PCR genotyping. Several plants were identified containing fixed WUS promoter indel mutations in the absence of the Cas9 transgene, suggesting a fixed promoter mutation in two generations (Figure 5j). After this line was generated, we set out to transform this promoter mutant line with a TDNA containing the PTA along with WUS sgRNA1 targeted the wild-type promoter sequence absent in the promoter mutant line. This line is currently being built. Identification of a line strongly expressing the PTA in the promoter mutant background will comprise the EGI *A. thaliana* line.

3.7 – Discussion of results and EGI future directions

Engineered Genetic Incompatibility is a next generation biocontainment strategy designed to keep transgenes from escaping into closely related weedy populations from a crop species. With this work we demonstrate a proof of concept for engineering EGI in the model plant *Arabidopsis thaliana*. Target genes were identified from the literature including the developmental regulators *wuschel* and *leafy cotyledon 1*. Initial attempts to engineer EGI utilized a TALE DNA binding domain fused to either the VP64 or VPR ADs. Transgenic plants were made expressing these PTAs targeted to the pro-apoptotic

gene Uba2b. While plants were identified with strong expression of the PTAs, we did not observe target gene activation or the expected phenotype of premature senescence. This is not necessarily a surprise, as many papers have shown direct fusion of a single AD to a DNA binding domain yields only modest gene activation.[159] It is likely these TALE-VP64 or TALE-VPR PTAs were not able to sufficiently activate the Uba2b target gene.

Next a transient agrobacterium delivery technique called Fast-TrACC was developed and used to target the dCas9-TV PTA to the EGI target gene *WUS*.[14,160] The dCas9-TV activator was initially discovered to be toxic in E. coli during the cloning process, likely due to the repetitive nature of the six TAL and two VP64 ADs fused end-to-end at the CTD of dCas9. The TV domain was redesigned using codon optimization to minimize direct repeat sequences and a potato IV2 intron was placed immediately upstream of the TV domain to stabilize the construct during cloning. This improved TV architecture, termed TV*, was used to activate the EGI target gene *WUS*. Fast-TrACC was performed with six different single sgRNAs, each targeting a unique region of the *WUS* core promoter. The sgRNAs ranged from 50bp to 1kb upstream of *WUS*, with the most active sgRNA being closest to the TSS. Transgenic plants were made to express the dCas9-TV* PTA alongside the three most active sgRNAs. Again, transgenic plants were identified highly expressing the PTAs but failed to activate the EGI target gene and produce the expected overexpression phenotype. Again, the most likely conclusion was that the dCas9-TV* PTA was not efficient enough at activating the target gene. From the transient agrobacterium assays we only observed modest ~15-fold gene activation of the WUS locus. Perhaps more robust activation in transient assays with a different PTA architecture would yield the expected overexpression result in transgenic plants.

To this end, the SunTag system was demonstrated in plants to be a strong activator.[60] A version of this SunTag system was generated capable of recruiting 24

copies of an activation domain to a single Cas9 molecule. Transient assays in A. thaliana protoplasts were carried out to compare the acivity of the SunTag system to dCas9-TV*. The EGI target genes WUS, LEC1, and STM were targeted for activation using a multi-guide array capable of co-expressing four sgRNAs targeting a single core promoter. The SunTag system outperformed dCas9-TV* by several orders of magnitude, reaching statistical significance across all three target genes. The improved plant-derived AD DREB2 produced higher gene activation across EGI target genes in comparison to the standard VP64 AD used by many groups. In addition, the expression of all three EGI target genes reached the same order of magnitude as the housekeeping gene PP2A with SunTag, indicating that transgenic plants expressing this PTA are more likely to yield overexpression phenotypes than other PTA architectures previously tested.

A crossing strategy was then set up to determine whether or not EGI target gene activation could produce a lethal phenotype. Because transgenes randomly integrate into the genome during agrobacterium mediated transformation, selection is applied by placing a resistance gene on the TDNA. Integrations of the transgene into euchromatin resulting in strong expression can survive when grown on selection media, while integrations with low expression of the transgene cannot survive. If a transgene is capable of driving lethality, it may be difficult to apply selection and differentiate between seedling lethality due to low transgene expression (selection) vs seedling lethality due to PTA-mediated EGI target gene activation. To overcome this, we separated selection and EGI target gene activation to occur in different generations. First, two different classes of transgenic lines were built; one line expressed the SunTag PTA along with a Hygromycin selectable marker, and the other lines expressed the sgRNAs along with a Kanamycin selectable marker. Transgenic lines were identified strongly expressing the PTA and sgRNA loci, as the visible reporters Firefly luciferase and GFP were also included on the TDNAs to track expression. Once strong expressing lines were

confirmed after two generations, crosses were performed between the PTA line and sgRNA lines for WUS and LEC1.

Hybrid seeds were collected from the crosses and placed on selection media for both Kanamycin and Hygromycin, ensuring strong expression of both transgenes. WUS hybrid seedlings failed to develop past the first set of true leaves, and GFP positive sectors of these seedlings showed hallmarks of WUS overexpression such as undifferentiated callous-like growths reminiscent of ectopic somatic embryogenesis.[128] The LEC1 hybrid seedlings failed to produce any root structures, and instead grew abnormal growths which protruded from the base of the plant where the roots were expected to form. Hybrid seedlings from each of these crosses were sacrificed for RNA isolation seven days post-germination, and gene expression was quantified. We detected strong PTA expression in all seedlings across both crosses, however, we observed variable expression of GFP which was linked to expression of the sgRNAs. The expression of this GFP was correlated with the observed target gene activation, particularly in the LEC1 hybrids.

Transgene silencing has been observed in many plants over the years.[161,162] In the particular case of these EGI crosses, hybrid offspring are undergoing two competing stressors. First, the seeds are growing on selection media. This process selects for strong expression of the transgene, or in other words prevents gene silencing as seedlings that silence the transgene would not be able to survive selection. Second, PTA-mediated gene activation of the EGI target gene is expected to perturb normal development and result in seedling lethality. Instances of transgene silencing in hybrid plants may be selected for, as plants that silence either transgene can then bypass the PTA-mediated lethality. The observed hybrid phenotypes and gene expression suggests plants may be silencing the sgRNA transgene, as indicated by low GFP expression particularly in the WUS crosses despite those same lines surviving selection one

generation earlier. This transgene silencing may be the result of cell populations that fail to silence the transgene early in development being outcompeted by cell populations that succeed in silencing the transgene only to die later due to growth on selection media.

These observations suggest that the use of tissue-specific promoters may be critical for ensuring target gene activation and lethal phenotype. Expressing the PTAs + sgRNAs constitutively resulted in potential silencing phenotypes. If the PTAs + sgRNAs are expressed at a certain developmental timepoint, transgene silencing may not occur due a reduced selective pressure to silence the transgene amongst cell populations during development because the PTAs are not expressed. Promoters like RPS5a and RPL23 are currently being cloned to drive expression of the PTA.[88] These promoters will drive strong mid to late embryonic expression of the PTA, as opposed to constitutive overexpression promoters like AtUbi10 or the CaMV 35S promoter.

Following observation of lethal phenotypes, promoter mutation lines needed to be created to protect the EGI line from targeting its own genome for lethal overexpression. To efficiently generate non-lethal mutations in the core promoter, it was preferred to use a single sgRNA. Studies have indicated that target a small genomic region with multiple sgRNAs can result in large deletions at high frequencies.[155,156] We wanted to avoid larger deletions in the core promoter region because deletion of the transcription start site can phenocopy a knockout mutation.[154] Protoplast transformation with each individual WUS sgRNA from the multi guide array indicated that sgRNA1 contributed the majority of observed gene activation. This allowed for promoter mutagenesis using a single sgRNA, resulting in either and A/T +1bp insertions in the seed region of the sgRNA. We then segregated away the Cas9 and retained the promoter mutation to comprise the EGI line. Currently we are transforming this promoter mutant with TDNAs expressing our optimized plant PTAs along with the WUS sgRNA1

targeting the non-mutated wild-type sequence. Once these lines are generated and high expression of the PTAs is confirmed, we will perform reciprocal crosses between the EGI xEGI, EGI x WT, and WT x WT. In this scenario, we anticipate the EGI line will be able to mate with itself, and the WT population can of course mate with itself. We hope to observe specific lethality in the EGI x WT crosses due to PTA-mediated EGI target gene overexpression, in this case activation of the endogenous WUS locus.

3.8 – Materials and Methods

Plant transformation and genotyping. To generate transgenic *A. thaliana* plants we performed floral dip protocols as previously described.[103] TDNA plasmids were constructed containing an antibiotic resistance gene along with the pFAST Oleosin-RFP transgene. *A. thaliana* ecotype Columbia-0 was grown to maturity and floral dipped with Agrobacterium strain GV3101 transformed with the described TDNAs. The antibiotic selection used in this study was either Kanamycin or Bialaphos resistance. T0 seeds were identified by antibiotic selection or screening for RFP+ seedlings. 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. Through each generation either GFP or Firefly luciferase was used to screen for plants expressing the transgene. For mutation analysis, plants were grown to maturity for ~3 weeks. Leaf tissue was collected for DNA extraction using the CTAB method.[163] PCR was performed across the sgRNA target site, followed by purification of the PCR product. This product was sent for Sanger sequencing, followed by Tracking of Indels by Decomposition (TIDE).[164] A commercial TIDE service was used from Synthego called ICE for submission of PCR amplicons for mutation analysis.

Fast-TrACC Fast-TrACC involves treating Agrobacterium cultures (GV3101) for 3 days prior to a 2 day co-culture with newly germinated seedlings. The first step is to grow the cultures overnight (8–12 h) in Luria broth (LB) with antibiotics [i.e., kanamycin (50 mg/mL) and gentamycin (50 mg/mL)] at 28°C. Next, cells are harvested by centrifugation and re-suspended to an OD_{600} of 0.3 in AB:MES200 salt solution (17.2 mM K_2HPO_4, 8.3 mM NaH_2PO_4, 18.7 mM NH_4Cl, 2 mM KCl, 1.25 mM $MgSO_4$, 100 µM $CaCl_2$, 10 µM $FeSO_4$, 50 mM MES, 2% glucose (w/v), 200 µM acetosyringone, pH 5.5) (Wu et al., 2014) and then grown overnight. The purpose of the AB:MES200 solution is to increase

the expression of *vir* genes. The culture is again centrifuged and resuspended to OD_{600} within the range of 0.10–0.18 (typically 0.14) in a 50:50 (v/v) mix of AB:MES200 salt solution and ½ MS liquid plant growth medium (1/2 MS salt supplemented with 0.5% sucrose (w/v), pH 5.5). Seeds are sterilized using 70% ethanol for 1 min and 50% bleach (v/v) (the hypochlorite concentration of the bleach was 7.4%) for 5 min. They are then rinsed 5 times with sterile water. Seeds are transferred to 6-well plates (~5 seeds per well in 2 mL ½ MS) and maintained in growth chambers (24°C, 16/8 h light/dark cycle). Individual species vary on their germination times (defined as initial cotyledon emergence) in liquid 12 MS: canola seedlings germinate in 2–3 days, *N. benthamiana* seedlings germinate in 3–4 days, tomatoes and potatoes germinate in ~7 days, peppers and eggplant germinate in ~14 days. Two days post germination, 12 MS media is removed and the treated Agrobacterium culture is added. The co-cultured seedlings are incubated for 2 days before being washed free of Agrobacterium using sterile water. The washed seedlings are returned to liquid ½ MS containing the antibiotic timentin at a concentration of 100 µM to effectively counter-select against residual Agrobacterium.

Firefly Luciferase Imaging Seedlings are analyzed for delivery of the T-DNA constructs containing a firefly luciferase reporter through long exposure imaging. Luciferin substrate (5 µL of 50 mM in ddH_2O stock into 2 mL of ½ MS, final concentration of 125 µM luciferin in 12 MS) is added to the ½ MS liquid culture with the seedlings to produce light. The plate of seedlings is then lightly shaken for 5 min to ensure proper mixing of the luciferin solution. Long-exposure imaging (5.5 min exposure using a UVP BioImaging Systems EpiChemi^3 Darkroom) is then performed to capture the luminescence.

RNA isolation and quantification. We isolate RNA from either protoplasts or leaf tissue according to 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. 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. The TRIzol protocol is followed 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. 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 55C and yield amplicon lengths be- tween 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.

Protoplast isolation and transformation. *S. viridis* and *A. thaliana* protoplasts were isolated and transformed as stated in previous protocols.[95,96] ME034 *S. viridis* plants are grown in a growth chamber set to 31C and 21C 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. Zea mays plants were grown in a growth chamber at 25C 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 midi prepped according to the manufacturer's protocol (Qiagen 12945) to ensure transformation grade endotoxin free DNA.

3.9 – Figure Legends

Figure 1 – Overview of engineered genetic incompatibility (EGI). **a** Macromolecular components that constitute programmable transcription factors (above), and schematic illustration showing lethal gene expression from a wild-type but not a refactored promoter (below) **b** Illustration of hybrid lethality upon mating of wild-type (orange cell) and EGI (green cell) parents. Macromolecular components are labeled in **a**, red DNA, signifies WT promoter, and blue DNA signifies refactored promoter. Skull and crossbones indicates a non-viable genotype. **c** Possible applications for engineered speciation.

Figure 2 – Testing the EGI target gene *Uba2b* with TALE PTAs. **a** Genomic target site of the EGI target gene Uba2b (AT2G41060). The core promoter region of this gene is shown in the dashed blue box, where TALE PTAs are targeted for gene activation. **b** TDNA backbone architecture used for floral dip into A. thaliana plants. The TALE DNA binding domain was engineered to target the indicated promoter regions in panel a. The VP64 or VPR AD was fused to the TALE DNA binding domain to comprise the PTA. **c** Image of T1 plants after 1-2 weeks on selection and moved to soil. No overexpression phenotypes were observed.

Figure 3 Testing the dCas9-TV* PTA using Fast-TrACC in *Arabidopsis thaliana*. **a** Cartoon images showing the dCas9-TV PTA. A refactored version of the TV domain was generated through codon optimization to reduce direct repeats along with placing the potato IV2 intron upstream of the TV domain for optimal expression in *E. coli*. **b** Gel images showing PCR amplification across the original TV domain (left) and the refactored TV* domain (right) in *E. coli*. Expected band size is 1.6kb. **c** Schematic of the Fast-TrACC workflow. Seedlings are grown on MS media in sterile culture. Agrobacterium tumefaciens is grown in the presence of AB:MES salts to induce *vir* gene expression. Seedlings are then co-cultured with the AB:MES induced agrobacterium.

After two days, seedlings are removed from co-culture and transferred to liquid media containing Timentin to kill off any residual agrobacterium. Downstream assays can then be performed such as quantification of gene expression. **d** Genomic region targeted for *WUS* overexpression. The core promoter was targeted with six different sgRNAs, ranging from 50bp to 1000bp upstream of the TSS. **e** Luciferase expression of A. thaliana seedlings 3 days post co-culture. Fast-TrACC was used to deliver TDNAs containing the dCas9-TV* PTA + sgRNA targeting the core promoter of *WUS*. **f** Gene activation of the *WUS* target in Fast-TrACC seedlings. Fold change was calculated using the delta delta Ct method comparing each sgRNA to the off-target sgRNA control. **g** Images of T1 transgenic plants generated with sgRNAs4-6 TDNAs used to activate WUS in panel f. Plants were selected on Kanamycin media to identify highly expressing PTA lines. No lethal phenotypes were observed.

Figure 4 EGI target gene activation by SunTag PTAs in *Arabidopsis thaliana* protoplasts. **a** Cartoon images of PTAs and sgRNA expression arrays used in this experiment. The dCas9-TV* PTA (left) described in Figure 3 was compared with the SunTag PTA (middle). Expression of sgRNAs from a multi-guide array was accomplished by flanking sgRNAs with tRNA motifs capable of self-processing, releasing four independent sgRNAs from a single transcript. **b** Three EGI target genes *WUS*, *LEC1*, and *STM* were targeted with multi-guide arrays for gene activation. Core promoter regions are shown along with sgRNA locations for each gene. **c** Results of protoplasts assays targeting the EGI targets shown in panel b. Expression is calculated relative to the housekeeping gene *PP2A*. SunTag PTAs using either the VP64 or DREB2 AD outperformed the dCas9-TV* PTA across all three target genes. **d-f** Same data as shown in panel c, but instead calculating Fold Change against a negative control for each individual target gene. SunTag PTAs for *WUS* (d) *LEC1* (e) and *STM* (f) reached statistical significance while dCas9-TV* did not reach statistical significance across all

three targets. Parametric t-tests were performed comparing each group against ST-VP64-NOG control in which no sgRNA was delivered. * = p<0.05, ** = p<0.01.

Figure 5 Testing and building the EGI *Arabidopsis thaliana* line**. a** TDNA constructs transformed to generate lines for crossing. The PTA line was generated using a TDNA lacking an sgRNA (top) while an sgRNA line was generated using a TDNA lacking the PTA (bottom). **b** The sgRNA lines from panel a targeted the *WUS* and *LEC1* core promoters at the indicated locations, ranging from 50 to 300bp upstream the annotated TSS. **c** Many PTA lines were selected and expression analysis was performed to detect expression of both the dCas9-24xGCN4 and scFv-VP64 transcripts. The dCas9 expression was always higher than scFv across all lines. **d** The PTA line contained a luciferase allowing for visualization of TDNA expression (left). The sgRNA line contained a GFP also allowing for visualization of TDNA expression (right). **e** Crosses were performed between a PTA line and a sgRNA line for either *LEC1* (left) or WUS (middle, right). *LEC1* progeny had a 50% survival rate after one week, while *WUS* progeny did not survive in either cross. **f** Images of hybrid seedlings three weeks post germination. *LEC1* hybrids did not develop roots, and instead GFP positive growths protruded. WUS hybrids did not develop past the first set of true leaves. GFP positive sectors could be seen displaying abnormal growths reminiscent of ectopic somatic callous-like growths. **g-h** One week post germination seedlings were sacrificed for RNA expression analysis. For *LEC1* hybrids two GFP positive and two GFP negative seedlings were selected for analysis (g). For *WUS* hybrids four seedlings were randomly selected from each of two independent crosses (h). Gene expression is shown relative to the housekeeping gene *PP2A*, and the colored dotted line indicates the basal expression level of the EGI target in the no sgRNA control (NOG). **i** *A. thaliana* protoplast assays to test the relative strength of each sgRNA in the *WUS* multi-guide array. The position of each sgRNA is noted (top) and the fold change in expression is shown relative to a no sgRNA (NOG)

control (bottom). **j** Promoter mutants were generated for *WUS* sgRNA1 using Cas9, with fixed +1bp A or T insertions detected across multiple T2 lines. The Cas9 transgene was segregated away prior to mutation analysis.

3.10 - Figures

Figure 1 – Overview of engineered genetic incompatibility (EGI)

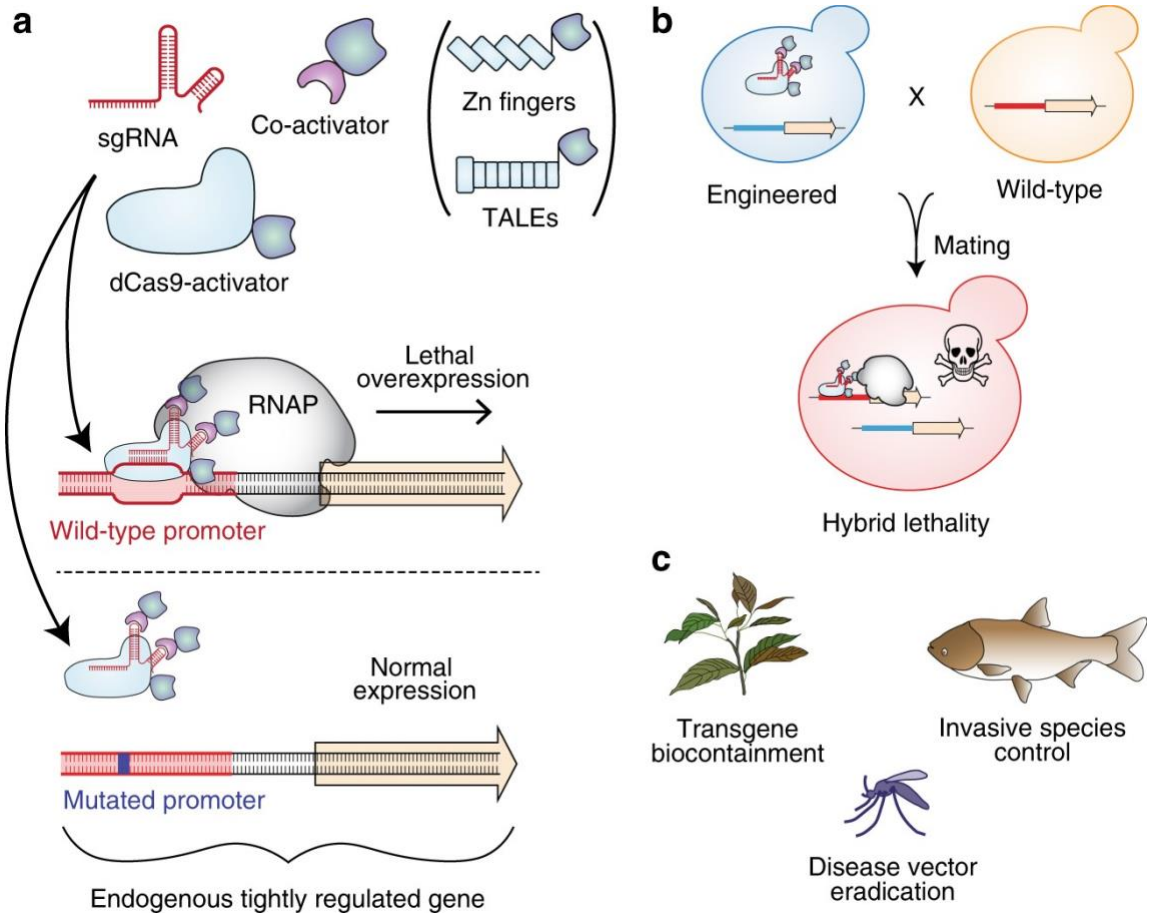


Figure 2 – Testing the EGI target gene *Uba2b* with TALE PTAs.

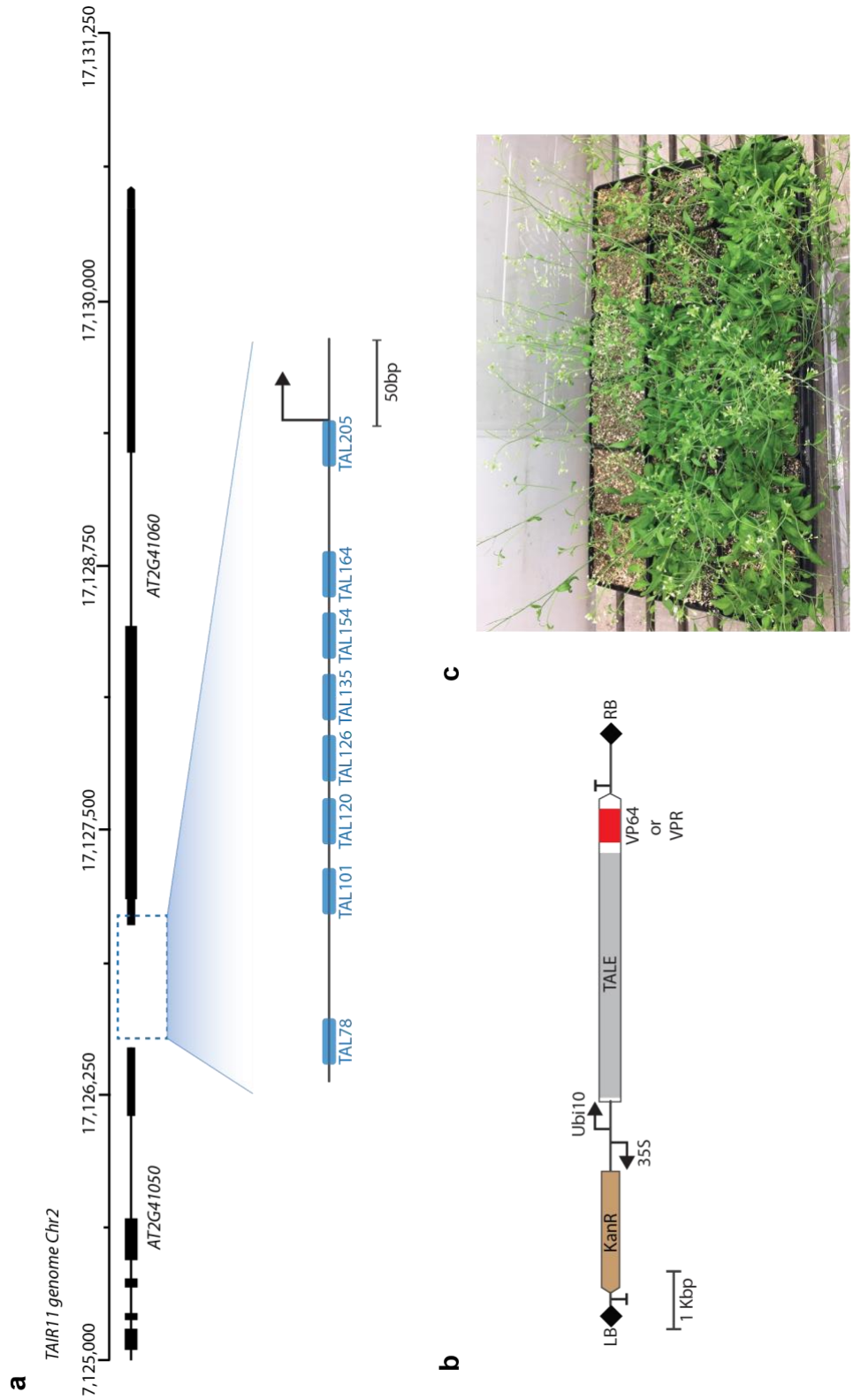


Figure 3 Testing the dCas9-TV* PTA using Fast-TrACC in *Arabidopsis thaliana*.

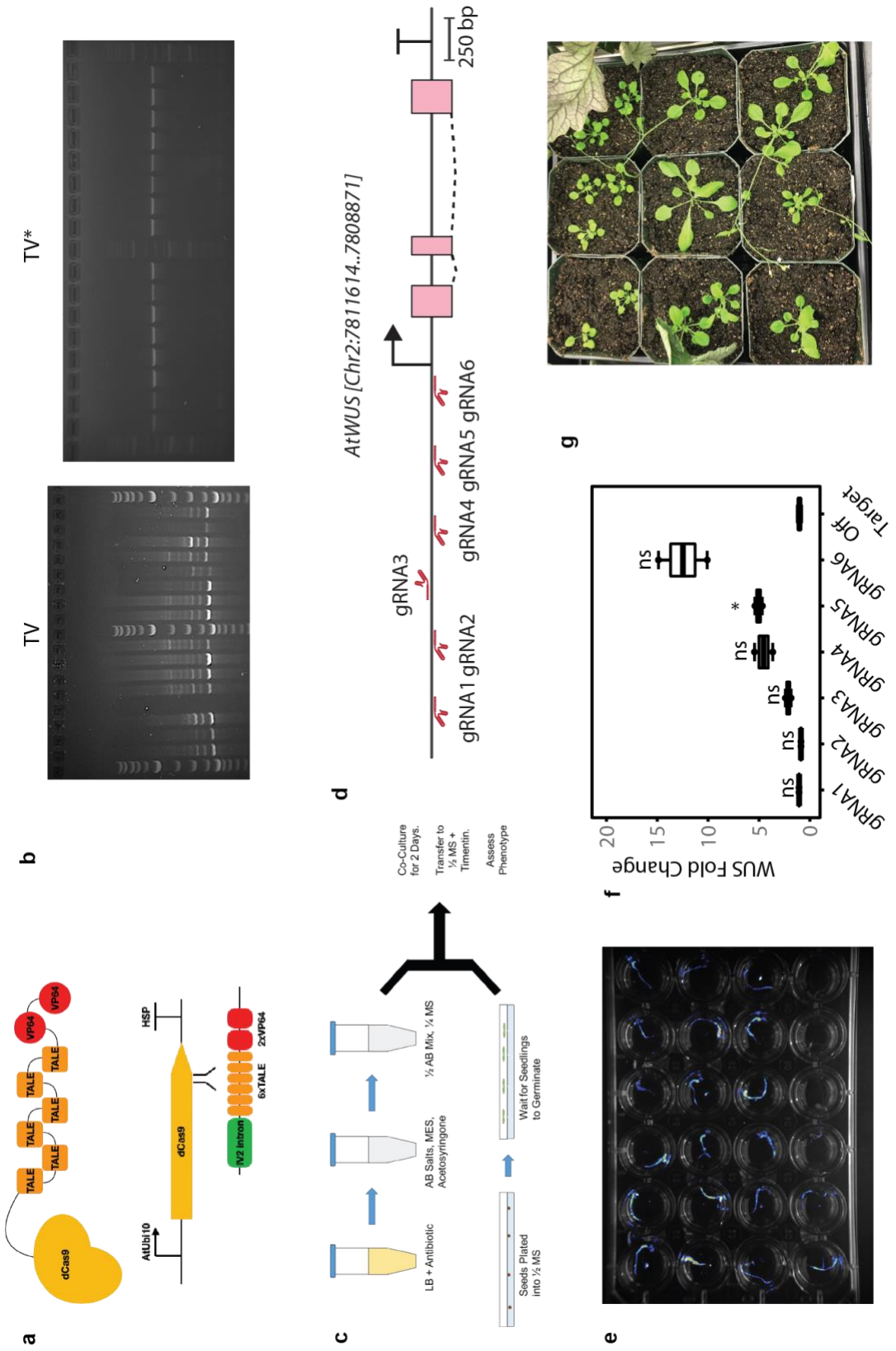


Figure 4 EGI target gene activation by SunTag PTAs in *Arabidopsis thaliana* protoplasts.

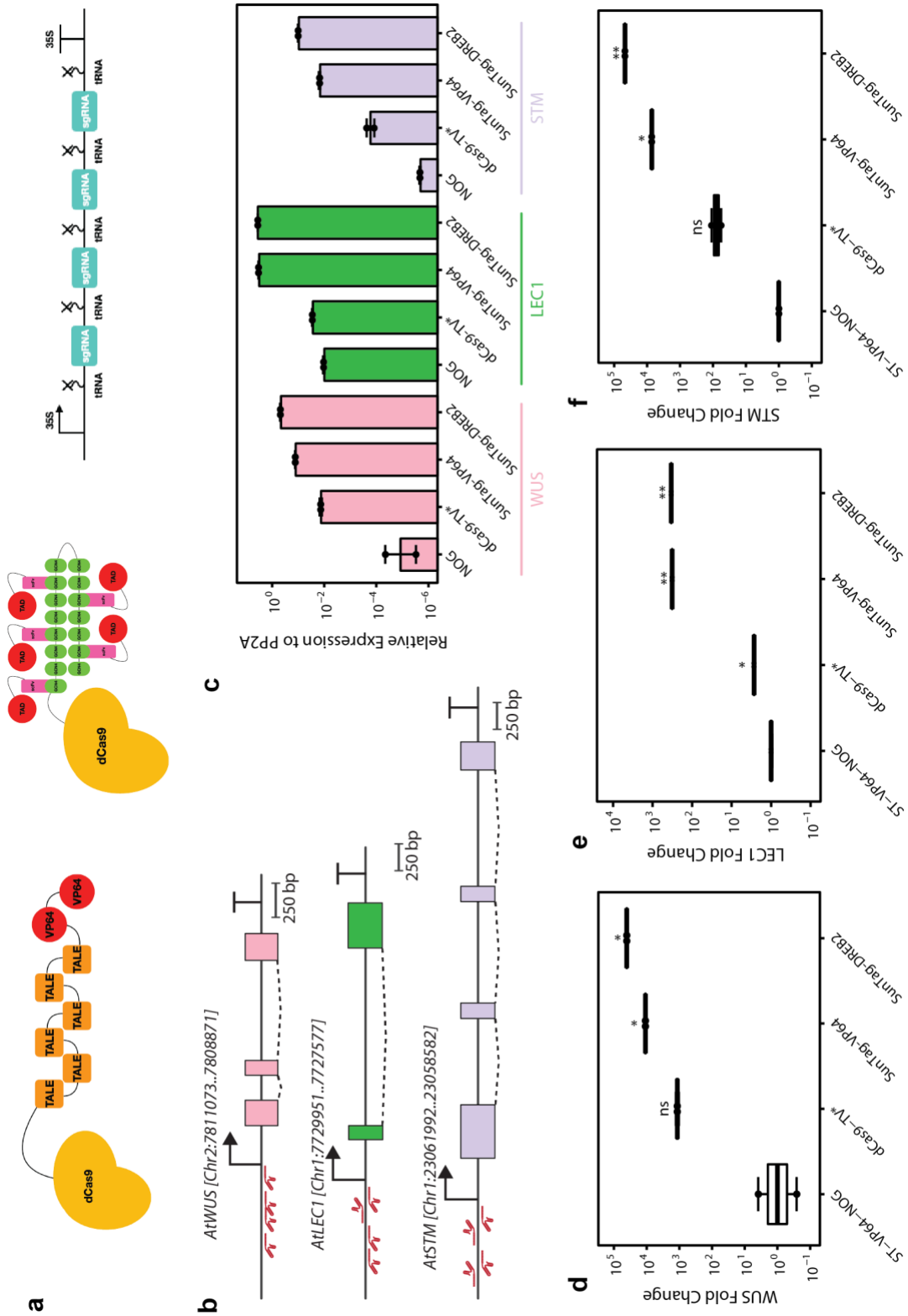
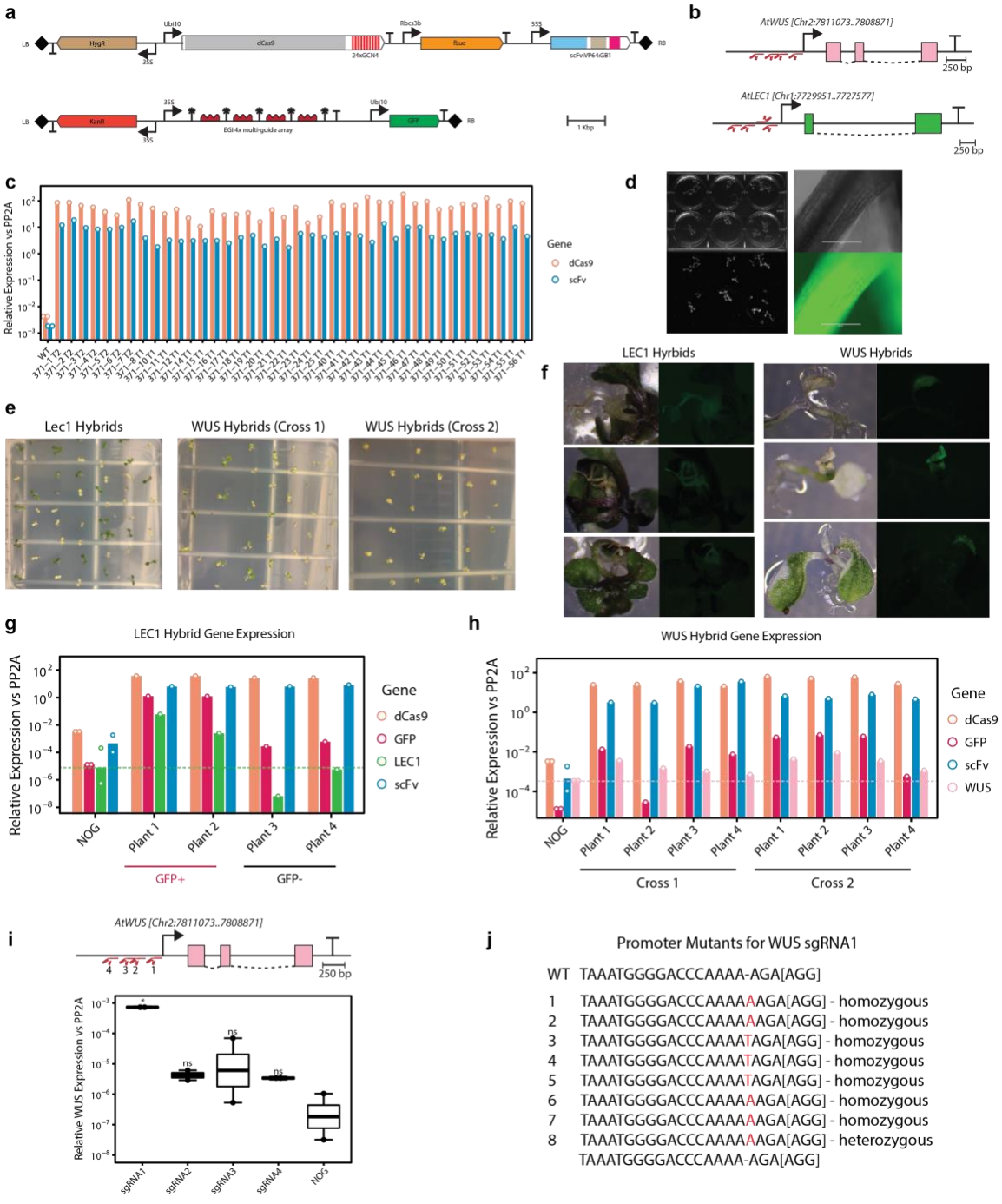


Figure 5 Testing and building the EGI *Arabidopsis thaliana* line



Chapter 4 – Future perspectives for engineering gene expression in plants using PTAs

When I first joined the Voytas and Smanski labs as a graduate student in 2017, I thought a lot about where agriculture was going and how gene editing could be used to improve crops. There are instances of single gene mutations generated by CRISPR-Cas9 that have improved crops, one example being waxy corn.[165] However, traits controlled by a single gene are rare and difficult to identify. For example 1671 single genes were each transformed into maize and tested in the field for their impact on traits like yield and drought tolerance.[166] Of those 1671, only 22 genes were identified as contributing to relevant traits.[166] This illustrates the difficulty of identifying and expressing single genes in transgenic plants to improve crops. Many agronomic traits of value today are likely controlled through the simultaneous action of many genes, and not a single gene. Complex traits like yield enhancement, drought tolerance, nitrogen use efficiency, plant height, plant lodging, and seed mass, among others, are currently being pursued for crop improvement with multiple genomic loci implicated in the underlying genetics of the traits.[166,167]

Transcription factors (TFs) were originally identified as good candidates for single modifications impacting complex traits in agriculture.[168] Nature has evolved transcription regulatory cascades to increase or decrease gene expression, impacting plant physiology to suit a given environment. At the crux of these cascades are transcription factors; proteins containing DNA binding domains and transcription activation or repression domains. Modifying the gene expression of TFs can result in differential expression of gene families involved in distinct biological processes, impacting complex traits like grain yield.[169]

To this end, building *de novo* transcription factors may allow for even more variety in the targetable gene families for activation or repression. Natural TFs are constrained in their target genes by the native DNA binding domain, natural expression

patterns, and post-translational regulatory mechanisms.[170] Conversely, artificial transcription factors (PTAs) can be expressed to target any region of the genome containing a protospacer-adjacent motif (PAM) and free from post-translational regulatory pathways. A key constraint with PTAs is choosing a promoter to drive the transcription factor, which is also a constraint when engineering native transcription factors in transgenic plants.[169] Thus, PTAs offer an alternative platform for engineering complex traits in plants through the coordinated activation or repression of target genes implicated in traits of interest.

These PTAs can target a broader range of promoter targets than natural TFs, with flexibility in the sgRNA binding site, or sites, for a given core promoter. Stacking multiple sgRNA targets in a single promoter can result in synergistic gene activation when targeted by PTAs.[26] This phenomena allows for the tuning of PTA-mediated gene expression by varying the number of sgRNA target sites in a given gene promoter. Currently multiple sgRNAs can be expressed using strategies such as tRNA or Csy4 arrays.[97] However, the maximum number of sgRNAs expressed in these systems is typically 6-8 sgRNAs per array.[97] To increase the number of targetable genes using a single PTA, improvements in multi-guide expression platforms will be critical.

In addition, different PTA programmable DNA binding domains have been demonstrated including different CRISPR-Cas systems along with TALE DNA binding domains.[20,30,159] These programmable DNA binding domains have different sequence constraints, with different CRISPR systems requiring different protospacer adjacent motifs (PAMs) or TALEs requiring a 5' T nucleotide for DNA targeting.[4,20,143] Thus, the number of targetable sites within a given core promoter can be increased by using different programmable DNA binding domains. Orthogonality can be also engineered by using different programmable DNA binding domains. Different CRISPR-Cas systems use different sgRNA scaffolds, meaning one can co-express two PTAs simultaneously.

This has been demonstrated using two variants of the Cas9 protein from either S. pyogenes and S. aureus.[171] Co-expression of two PTAs each with different DNA binding domains would allow for co-expression of two artificial TFs, each capable of independently regulating different sets of genes to impact traits of interest.

The work described in Chapter Two resulted in improvements to plant PTAs. Of particular interest were ADs capable of increasing transcription activation to levels higher than the oft-used VP64. While a handful of domains were identified that were better than VP64 in plant cells, we also identified several other ADs that function equal to or slightly worse than VP64. These ADs can all be deployed to achieve different levels of gene activation. We anticipate these tools can be used by plant biologists to activate target genes of interest in plants. We hope that plant-optimized PTAs will be used to activate families of genes by to co-expression of multiple sgRNAs from a single TDNA. This would allow for hypothesis testing through the engineering of artificial transcription factors in plants targeting sets of genes implicated in a trait of interest.

Ultimately, agriculture must address many potential hurdles in the coming years. World populations continue to grow and are expected to reach 9.7 billion by 2050. Global climate change is resulting in environmental changes to which agriculture must react. To produce an efficient, sustainable and resilient food supply, crop improvement will continue to be critical. Traditionally natural variation has been harnessed for years to improve crops through selective breeding. Underpinning much of this variation within species is differential gene expression.[172–174] As we move into the gene editing era of crop improvement, manipulating plant genomes has become ever more feasible. Coding sequence mutations have been generated to impact traits of interest.[165] Cis regulatory elements have been mutated, generating new expression variants impacting a trait.[175] Finally, PTAs have been used to activate target genes of interest and impact phenotypes.[176] PTAs are an important tool in the CRISPR toolbox that will undoubtedly

be critical for controlling gene expression in plants. Crop improvement over the next 50 years will require modifications to gene expression and plant-optimized PTAs provide a method to efficiently activate gene expression across crops.

Bibliography

1. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, (2018).
2. Garneau, J. E. *et al.* The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
3. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).
4. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* **337**, 816–821 (2012).
5. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
6. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
7. Green, M. R. Eukaryotic transcription activation: Right on target. *Mol. Cell* **18**, 399–402 (2005).
8. Petrenko, N., Jin, Y., Dong, L., Wong, K. H. & Struhl, K. Requirements for rna polymerase ii preinitiation complex formation in vivo. *Elife* **8**, (2019).
9. Allen, B. L. & Taatjes, D. J. The Mediator complex: A central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 155–166 (2015).
10. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e16 (2018).

11. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. GAL4-VP16 is an unusually potent transcriptional activator. *Nat. 1988 3356190* **335**, 563–564 (1988).

12. Beerli, R. R., Segal, D. J., Dreier, B. & Barbas, C. F. 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. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14628–14633 (1998).

13. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods 2015 124* **12**, 326–328 (2015).

14. Li, Z. *et al.* A potent Cas9-derived gene activator for plant and mammalian cells. *Nat. Plants 2017 312* **3**, 930–936 (2017).

15. Schmitz, M. L. & Baeuerle, P. A. The p65 subunit is responsible for the strong transcription activating potential of NF-kappa B. *EMBO J.* **10**, 3805–3817 (1991).

16. Hardwick, J. M., Tse, L., Applegren, N., Nicholas, J. & Veliuona', M. A. The Epstein-Barr virus R transactivator (Rta) contains a complex, potent activation domain with properties different from those of VP16. *J. Virol.* **66**, 5500 (1992).

17. Green, M., Schuetz, T. J., Sullivan, E. K. & Kingston, R. E. A heat shock-responsive domain of human HSF1 that regulates transcription activation domain function. *Mol. Cell. Biol.* **15**, 3354–3362 (1995).

18. Stockinger, E. J., Gilmour, S. J. & Thomashow, M. F. Arabidopsis thaliana CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1035–1040 (1997).

19. Tiwari, S. B. *et al.* The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant J.* **70**, 855–865 (2012).

20. Young, J. K. *et al.* The repurposing of type I-E CRISPR-Cascade for gene

activation in plants. *Commun. Biol.* **2**, (2019).

21. Zhu, W., Yang, B., Chittoor, J. M., Johnson, L. B. & White, F. F. AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. *Mol. Plant-Microbe Interact.* **11**, 824–832 (1998).
22. Sigler, P. B. Acid blobs and negative noodles. *Nature* **333**, 210–212 (1988).
23. Babu, M. M. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* **44**, 1185–1200 (2016).
24. Kwon, I. *et al.* XPhosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049 (2013).
25. Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).
26. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23**, 1163–1171 (2013).
27. Farzadfard, F., Perli, S. D. & Lu, T. K. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth. Biol.* **2**, 604–613 (2013).
28. Gilbert, L. A. *et al.* XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442 (2013).
29. Kearns, N. A. *et al.* Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Dev.* **141**, 219–223 (2014).
30. Maeder, M. L. *et al.* CRISPR RNA–guided activation of endogenous human genes. *Nat. Methods 2013 1010* **10**, 977–979 (2013).
31. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR- Cas9–based transcription factors. *Nat. Methods* **10**, (2013).
32. Piatek, A. *et al.* RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors. *Plant Biotechnol. J.* **13**, 578–589 (2015).

33. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).

34. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).

35. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **159**, 635–646 (2014).

36. Kunii, A. *et al.* Three-Component Repurposed Technology for Enhanced Expression: Highly Accumulable Transcriptional Activators via Branched Tag Arrays. *Cris. J.* **1**, 337–347 (2018).

37. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759–771 (2015).

38. Zhang, X. *et al.* Multiplex gene regulation by CRISPR-ddCpf1. *Cell Discov.* **3**, (2017).

39. Tak, Y. E. *et al.* Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. *Nat. Methods* **14**, 1163–1166 (2017).

40. Zhang, X. *et al.* Genetic editing and interrogation with Cpf1 and caged truncated pre-tRNA-like crRNA in mammalian cells. *Cell Discov.* **4**, (2018).

41. Zhang, X. *et al.* Gene activation in human cells using CRISPR/Cpf1-p300 and CRISPR/Cpf1-SunTag systems. *Protein Cell* **9**, 380–383 (2018).

42. Nihongaki, Y., Otabe, T., Ueda, Y. & Sato, M. A split CRISPR–Cpf1 platform for inducible genome editing and gene activation. *Nat. Chem. Biol.* **15**, 882–888 (2019).

43. Kim, D. *et al.* Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868 (2016).

44. Kleinstiver, B. P. *et al.* Genome-wide specificities of CRISPR-Cas Cpf1 nucleases

in human cells. *Nat. Biotechnol.* **34**, 869–874 (2016).

45. Singh, D. *et al.* Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5444–5449 (2018).

46. Zetsche, B. *et al.* Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).

47. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).

48. Anderson, G. M. & Freytag, S. O. Synergistic activation of a human promoter in vivo by transcription factor Sp1. *Mol. Cell. Biol.* **11**, 1935–1943 (1991).

49. Carey, M., Lin, Y. S., Green, M. R. & Ptashne, M. A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature* **345**, 361–364 (1990).

50. Pettersson, M. & Schaffner, W. Synergistic activation of transcription by multiple binding sites for NF-κB even in absence of co-operative factor binding to DNA. *J. Mol. Biol.* **214**, 373–380 (1990).

51. Perez-Pinera, P. *et al.* Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat. Methods* **10**, (2013).

52. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647–661 (2014).

53. Hu, J. *et al.* Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic Acids Res.* **42**, 4375–4390 (2014).

54. Deaner, M. & Alper, H. S. Systematic testing of enzyme perturbation sensitivities via graded dCas9 modulation in Saccharomyces cerevisiae. *Metab. Eng.* **40**, 14–22 (2017).

55. Gong, X., Zhang, T., Xing, J., Wang, R. & Zhao, Y. Positional effects on efficiency

of CRISPR/Cas9-based transcriptional activation in rice plants. *aBIOTECH* **1**, (2020).

56. Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat. Methods* **13**, 563–567 (2016).

57. Lin, S., Ewen-Campen, B., Ni, X., Housden, B. E. & Perrimon, N. In vivo transcriptional activation using CRISPR/Cas9 in Drosophila. *Genetics* **201**, 433–442 (2015).

58. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).

59. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).

60. Papikian, A., Liu, W., Gallego-Bartolomé, J. & Jacobsen, S. E. Site-specific manipulation of Arabidopsis loci using CRISPR-Cas9 SunTag systems. *Nat. Commun. 2019 101* **10**, 1–11 (2019).

61. Norman, T. M. *et al.* Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science (80-.).* **365**, 786–793 (2019).

62. Liu, Y. *et al.* CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. *Cell Stem Cell* **23**, 758-771.e8 (2018).

63. Boettcher, M. *et al.* Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat. Biotechnol.* **36**, 170–178 (2018).

64. Yang, J. *et al.* Genome-Scale CRISPRa Screen Identifies Novel Factors for Cellular Reprogramming. *Stem Cell Reports* **12**, 757–771 (2019).

65. Joung, J. *et al.* Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* **548**, 343–346 (2017).

66. Chong, Z. S., Ohnishi, S., Yusa, K. & Wright, G. J. Pooled extracellular receptor-

ligand interaction screening using CRISPR activation. *Genome Biol.* **19**, (2018).

67. Ramkumar, P., Kampmann, M. & Qian, C. CRISPR-based genetic interaction maps inform therapeutic strategies in cancer. *Transl. Cancer Res.* **7**, S61–S67 (2018).

68. Horlbeck, M. A. *et al.* Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* **5**, (2016).

69. Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, (2018).

70. Brauna, C. J. *et al.* Versatile in vivo regulation of tumor phenotypes by dCas9-mediated transcriptional perturbation. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3892–E3900 (2016).

71. Wangensteen, K. J. *et al.* Combinatorial genetics in liver repopulation and carcinogenesis with a in vivo CRISPR activation platform. *Hepatology* **68**, 663–676 (2018).

72. Zhou, H. *et al.* In vivo simultaneous transcriptional activation of multiple genes in the brain using CRISPR–dCas9-activator transgenic mice. *Nat. Neurosci. 2018 213* **21**, 440–446 (2018).

73. Schoger, E. *et al.* CRISPR-Mediated Activation of Endogenous Gene Expression in the Postnatal Heart. *Circ. Res.* 6–24 (2020). doi:10.1161/CIRCRESAHA.118.314522

74. Wang, G. *et al.* Multiplexed activation of endogenous genes by CRISPRa elicits potent antitumor immunity. *Nat. Immunol.* **20**, 1494–1505 (2019).

75. Hsu, M. N. *et al.* Coactivation of Endogenous Wnt10b and Foxc2 by CRISPR Activation Enhances BMSC Osteogenesis and Promotes Calvarial Bone Regeneration. *Mol. Ther.* **28**, 441–451 (2020).

76. Matharu, N. *et al.* CRISPR-mediated activation of a promoter or enhancer rescues

obesity caused by haploinsufficiency. *Science (80-.).* **363**, (2019).

77. Ewen-Campen, B. *et al.* Optimized strategy for in vivo Cas9-activation in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9409–9414 (2017).

78. Jia, Y. *et al.* Next-generation CRISPR/Cas9 transcriptional activation in Drosophila using flySAM. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 4719–4724 (2018).

79. Zirin, J. *et al.* Large-scale transgenic Drosophila resource collections for loss- And gain-of-function studies. *Genetics* **214**, 755–767 (2020).

80. Long, L. *et al.* Regulation of transcriptionally active genes via the catalytically inactive Cas9 in C. elegans and D. rerio. *Cell Res.* **25**, 638–641 (2015).

81. Lowder, L. G. *et al.* A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation. *Plant Physiol.* **169**, 971–85 (2015).

82. Park, J. J., Dempewolf, E., Zhang, W. & Wang, Z. Y. RNA-guided transcriptional activation via CRISPR/dCas9 mimics overexpression phenotypes in Arabidopsis. *PLoS One* **12**, (2017).

83. Lee, J. E., Neumann, M., Duro, D. I. & Schmid, M. CRISPR-based tools for targeted transcriptional and epigenetic regulation in plants. *PLoS One* **14**, e0222778 (2019).

84. Lowder, L. G. *et al.* Robust Transcriptional Activation in Plants Using Multiplexed CRISPR-Act2.0 and mTALE-Act Systems. *Mol. Plant* **11**, 245–256 (2018).

85. Jensen, E. D. *et al.* Transcriptional reprogramming in yeast using dCas9 and combinatorial gRNA strategies. *Microb. Cell Fact.* **16**, (2017).

86. Lian, J., Hamedirad, M., Hu, S. & Zhao, H. Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. *Nat. Commun.* **8**, (2017).

87. Maselko, M. *et al.* Engineering multiple species-like genetic incompatibilities in insects. *Nat. Commun.* **11**, (2020).

88. Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D. & Penin, A.

A. A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant J.* **88**, 1058–1070 (2016).

89. Knauer, S. *et al.* A high-resolution gene expression atlas links dedicated meristem genes to key architectural traits. *Genome Res.* **29**, 1962–1973 (2019).
90. Pang, Y. *et al.* High-Resolution Genome-wide Association Study Identifies Genomic Regions and Candidate Genes for Important Agronomic Traits in Wheat. *Mol. Plant* **13**, 1311–1327 (2020).
91. He, F. *et al.* Genomic variants affecting homoeologous gene expression dosage contribute to agronomic trait variation in allopolyploid wheat. *Nat. Commun. 2022 131* **13**, 1–15 (2022).
92. Armando Casas-Mollano, J., Zinselmeier, M. H., Erickson, S. E. & Smanski, M. J. CRISPR-Cas Activators for Engineering Gene Expression in Higher Eukaryotes. *Cris. J.* **3**, 350–364 (2020).
93. Chiarella, A. M. *et al.* Dose-dependent activation of gene expression is achieved using CRISPR and small molecules that recruit endogenous chromatin machinery. *Nat. Biotechnol.* **38**, 50 (2020).
94. Upadhyay, A. *et al.* Genetically engineered insects with sex-selection and genetic incompatibility enable population suppression. *Elife* **11**, (2022).
95. Yoo, S. D., Cho, Y. H. & Sheen, J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc. 2007 27* **2**, 1565–1572 (2007).
96. Sychla, A., Casas-Mollano, J. A., Zinselmeier, M. H. & Smanski, M. Characterization of Programmable Transcription Activators in the Model Monocot Setaria viridis Via Protoplast Transfection. in *Protoplast Technology, Methods in Molecular Biology* (eds. Wang, K. & Zhang Feng) **2464**, 223–244 (Humana, 2022).

97. Čermák, T. *et al.* A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. *Plant Cell* **29**, 1196–1217 (2017).

98. Sherf, B. A., Navarro, S. L., Hannah, R. R. & Wood, K. V. Dual-Luciferase TM Reporter Assay: An Advanced Co-Reporter Technology Integrating Firefly and Renilla Luciferase Assays. *Promega Notes Mag. Number* **57**, 2 (1996).

99. Luehrsen, K. R., de Wet, J. R. & Walbot, V. [35] Transient expression analysis in plants using firefly luciferase reporter gene. *Methods Enzymol.* **216**, 397–414 (1992).

100. Benfey, P. N. & Chua, N. H. The Cauliflower Mosaic Virus 35S Promoter: Combinatorial Regulation of Transcription in Plants. *Science (80-.).* **250**, 959–966 (1990).

101. Casas-Mollano, J. A., Zinselmeier, M. H. & Smanski, M. J. Efficient gene activation in plants by the MoonTag programmable transcriptional activator. *Manuscr. Prep.* (2022).

102. Boersma, S. *et al.* Multi-Color Single-Molecule Imaging Uncovers Extensive Heterogeneity in mRNA Decoding. *Cell* **178**, 458-472.e19 (2019).

103. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium -mediated transformation of Arabidopsis thaliana. *Plant J.* **16**, 735–743 (1998).

104. Shimada, T. L., Shimada, T. & Hara-Nishimura, I. A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana. *Plant J.* **61**, 519–528 (2010).

105. Takada, S. & Goto, K. 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. *Plant Cell* **15**, 2856–2865 (2003).

106. Krzymuski, M. *et al.* The dynamics of FLOWERING LOCUS T expression encodes long-day information. *Plant J.* **83**, 952–961 (2015).

107. Tak, Y. E. *et al.* Augmenting and directing long-range CRISPR-mediated activation in human cells. *Nat. Methods 2021 189* **18**, 1075–1081 (2021).

108. Zicola, J., Liu, L., Tänzler, P. & Turck, F. Targeted DNA methylation represses two enhancers of FLOWERING LOCUS T in Arabidopsis thaliana. *Nat. Plants 2019 53* **5**, 300–307 (2019).

109. Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebley, J. Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet. 2011 4311* **43**, 1160–1163 (2011).

110. Ricci, W. A. *et al.* Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants 2019 512* **5**, 1237–1249 (2019).

111. Fisher, A. C. & DeLisa, M. P. Efficient isolation of soluble intracellular single-chain antibodies using the twin-arginine translocation machinery. *J. Mol. Biol.* **385**, 299 (2009).

112. Kronqvist, N. *et al.* Efficient protein production inspired by how spiders make silk. *Nat. Commun. 2017 81* **8**, 1–15 (2017).

113. Oerke, E.-C. Crop losses to pests. *Jounrla Agric. Sci.* **144**, 31–43 (2006).

114. Zhang, Z. P. Development of chemical weed control and integrated weed management in China. *Weed Biol. Manag.* **3**, 197–203 (2003).

115. Livingston, M. *et al.* Economic Research Service Economic Research Report Number 184 The Economics of Glyphosate Resistance Management in Corn and Soybean Production. (2015).

116. Green, J. M. The rise and future of glyphosate and glyphosate-resistant crops. *Pest Manag. Sci.* **74**, 1035–1039 (2018).

117. Heap, I. & Duke, S. O. Overview of glyphosate-resistant weeds worldwide. *Pest*

Manag. Sci. **74**, 1040–1049 (2018).

118. Boerboom, C. GWC-1 The Glyphosate, Weeds, and Crops Series Facts About Glyphosate-Resistant Weeds The Glyphosate, Weeds, and Crops Series.

119. Arriola, P. E. & Ellstrand, N. C. Crop-To-Weed Gene Flow in the Genus Sorghum (Poaceae): Spontaneous Interspecific Hybridization between Johnsongrass , Sorghum halepense , and Crop Sorghum , S . Bicolor. *Am. J. Bot.* **83**, 1153–1159 (1996).

120. Gressel, J. Dealing with transgene flow of crop protection traits from crops to their relatives. *Pest Manag. Sci.* **71**, 658–667 (2015).

121. Gealy, D. Gene Movement between Rice (Oryza sativa) and Weedy Rice (Oryza sativa) — a U.S. Temperate Rice Perspective. *Crop Feral. Volunt.* 323–354 (2005). doi:10.1201/9781420037999.CH20

122. Clark, M. & Maselko, M. Transgene Biocontainment Strategies for Molecular Farming. *Front. Plant Sci.* **11**, 210 (2020).

123. Kuvshinov, V., Anisimov, A., Yahya, B. M. & Kanerva, A. Double recoverable block of function--a molecular control of transgene flow with enhanced reliability. *Environ. Biosafety Res.* **4**, 103–112 (2005).

124. Lombardo, L. Genetic use restriction technologies: a review. *Plant Biotechnol. J.* **12**, 995–1005 (2014).

125. Maselko, M., Heinsch, S. & Smanski, M. Engineering Synthetic Barriers to Sexual Reproduction. **40**, 7536 (2016).

126. Maselko, M., Heinsch, S. C., Chacón, J. M., Harcombe, W. R. & Smanski, M. J. Engineering species-like barriers to sexual reproduction. *Nat. Commun. 2017 81* **8**, 1–7 (2017).

127. Klingseisen, A., Clark, I. B. N., Gryzik, T. & Müller, H. A. Differential and overlapping functions of two closely related Drosophila FGF8-like growth factors

in mesoderm development. *Development* **136**, 2393–2402 (2009).

128. Zuo, J., Niu, Q.-W., Frugis, G. & Chua, N.-H. The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. *Plant J.* **30**, 349–359 (2002).

129. Lotan, T. *et al.* Arabidopsis LEAFY COTYLEDON1 Is Sufficient to Induce Embryo Development in Vegetative Cells. *Cell* **93**, 1195–1205 (1998).

130. Zuo, J., Niu, Q. W., Frugis, G. & Chua, N. H. The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. *Plant J.* **30**, 349–359 (2002).

131. Gallois, J. L., Woodward, C., Reddy, G. V. & Sablowski, R. Combined SHOOT MERISTEMLESS and WUSCHEL trigger ectopic organogenesis in Arabidopsis. *Development* **129**, 3207–3217 (2002).

132. Kim, C. Y., Bove, J. & Assmann, S. M. Overexpression of wound-responsive RNA-binding proteins induces leaf senescence and hypersensitive-like cell death. *New Phytol.* **180**, 57–70 (2008).

133. Lin, R. & Wang, H. Targeting proteins for degradation by Arabidopsis COP1: Teamwork is what matters. *J. Integr. Plant Biol.* **49**, 35–42 (2007).

134. Osterlund, M. T. & Deng, X. W. Multiple photoreceptors mediate the light-induced reduction of GUS-COP1 from Arabidopsis hypocotyl nuclei. *Plant J.* **16**, 201–208 (1998).

135. Cluis, C. P., Mouchel, C. F. & Hardtke, C. S. The Arabidopsis transcription factor HY5 integrates light and hormone signaling pathways. *Plant J.* **38**, 332–347 (2004).

136. Shin, B. *et al.* AtMYB21, a gene encoding a flower-specific transcription factor, is regulated by COP1. *Plant J.* **30**, 23–32 (2002).

137. Eide, L. *et al.* Comprehensive proteomic analysis of the human spliceosome. *Nat. 2002 4196903* **419**, 182–185 (2002).

138. Graveley, B. R. Sorting out the complexity of SR protein functions. *RNA* **6**, 1197–

1211 (2000).

139. Smith, C. W. J. & Valcárcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25**, 381–388 (2000).

140. Lazar, G. & Goodman, H. M. The Arabidopsis splicing factor SR1 is regulated by alternative splicing. *Plant Mol. Biol. 2000 424* **42**, 571–581 (2000).

141. Kalyna, M., Lopato, S. & Barta, A. Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. *Mol. Biol. Cell* **14**, 3565–77 (2003).

142. Voytas, D. F. & Joung, J. K. DNA binding made easy. *Science (80-.).* **326**, 1491–1492 (2009).

143. Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science (80-.).* **326**, 1501 (2009).

144. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting (Nucleic Acids Research (2011) 39 (e82) DOI: 10.1093/nar/gkr218). *Nucleic Acids Res.* **39**, 7879 (2011).

145. Christian, M. *et al.* Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* **186**, 757–761 (2010).

146. Christian, M., Qi, Y., Zhang, Y. & Voytas, D. F. Targeted mutagenesis of Arabidopsis thaliana using engineered TAL effector nucleases. *G3 (Bethesda).* **3**, 1697–705 (2013).

147. Altpeter, F. *et al.* Advancing crop transformation in the era of genome editing. *Plant Cell* **28**, 1510–1520 (2016).

148. Lin, C. S. *et al.* Application of protoplast technology to CRISPR/Cas9 mutagenesis: from single-cell mutation detection to mutant plant regeneration. *Plant Biotechnol. J.* **16**, 1295–1310 (2018).

149. Janssen, B. J. & Gardner, R. C. Localized transient expression of GUS in leaf

discs following cocultivation with Agrobacterium. *Plant Mol. Biol.* **14**, 61–72 (1990).

150. Ali, Z., Eid, A., Ali, S. & Mahfouz, M. M. Pea early-browning virus-mediated genome editing via the CRISPR/Cas9 system in Nicotiana benthamiana and Arabidopsis. *Virus Res.* **244**, 333–337 (2018).
151. Wu, H. Y. *et al.* AGROBEST: An efficient Agrobacterium-mediated transient expression method for versatile gene function analyses in Arabidopsis seedlings. *Plant Methods* **10**, (2014).
152. Maher, M. F. *et al.* Plant gene editing through de novo induction of meristems. *Nat. Biotechnol.* **38**, 84–89 (2020).
153. Wang, Y. *et al.* Deletion of a target gene in Indica rice via CRISPR/Cas9. *Plant Cell Rep.* **36**, 1333–1343 (2017).
154. Bookstein, R. *et al.* Promoter deletion and loss of retinoblastoma gene expression in human prostate carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 7762–7766 (1990).
155. Zhou, J. *et al.* Dual sgRNAs facilitate CRISPR/Cas9-mediated mouse genome targeting. *FEBS J.* **281**, 1717–1725 (2014).
156. Pauwels, L. *et al.* A Dual sgRNA Approach for Functional Genomics in Arabidopsis thaliana. *G3 Genes|Genomes|Genetics* **8**, 2603–2615 (2018).
157. Taylor, M. S. *et al.* Heterotachy in Mammalian Promoter Evolution. *PLOS Genet.* **2**, e30 (2006).
158. Cox, K. L. *et al.* TAL effector driven induction of a SWEET gene confers susceptibility to bacterial blight of cotton. *Nat. Commun. 2017 81* **8**, 1–14 (2017).
159. Lowder, L. G. *et al.* Robust Transcriptional Activation in Plants Using Multiplexed CRISPR-Act2.0 and mTALE-Act Systems. *Mol. Plant* **11**, 245–256 (2018).
160. Nasti, R. A., Zinselmeier, M. H., Vollbrecht, M., Maher, M. F. & Voytas, D. F. Fast-

TrACC: A Rapid Method for Delivering and Testing Gene Editing Reagents in Somatic Plant Cells. *Front. Genome Ed.* **0**, 32 (2021).

161. Beclin, C. *et al.* Transgene-induced gene silencing in plants Cite this paper Transgene-induced gene silencing in plants. *Plant J.* 651–659 (1998).
162. Iyer, L. M., Kumpatla, S. P., Chandrasekharan, M. B. & Hall, T. C. Transgene silencing in monocots. *Plant Mol. Biol. 2000 432* **43**, 323–346 (2000).
163. Porebski, S., Bailey, L. G. & Baum, B. R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Report. 1997 151* **15**, 8–15 (1997).
164. Brinkman, E. K., Chen, T., Amendola, M. & Van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, (2014).
165. Gao, H. *et al.* Superior field performance of waxy corn engineered using CRISPR–Cas9. *Nat. Biotechnol. 2020 385* **38**, 579–581 (2020).
166. Simmons, C. R. *et al.* Successes and insights of an industry biotech program to enhance maize agronomic traits. *Plant Sci.* **307**, 110899 (2021).
167. Diers, B. W. *et al.* Genetic Architecture of Soybean Yield and Agronomic Traits. *G3 Genes|Genomes|Genetics* **8**, 3367–3375 (2018).
168. Century, K., Reuber, T. L. & Ratcliffe, O. J. Regulating the Regulators: The Future Prospects for Transcription-Factor-Based Agricultural Biotechnology Products. *Plant Physiol.* **147**, 20–29 (2008).
169. Wu, J. *et al.* Overexpression of zmm28 increases maize grain yield in the field. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23850–23858 (2019).
170. Ishihama, N. & Yoshioka, H. Post-translational regulation of WRKY transcription factors in plant immunity. *Curr. Opin. Plant Biol.* **15**, 431–437 (2012).
171. Chen, B. *et al.* Expanding the CRISPR imaging toolset with Staphylococcus

aureus Cas9 for simultaneous imaging of multiple genomic loci. *Nucleic Acids Res.* **44**, e75–e75 (2016).

172. Lasky, J. R. *et al.* Natural Variation in Abiotic Stress Responsive Gene Expression and Local Adaptation to Climate in Arabidopsis thaliana. *Mol. Biol. Evol.* **31**, 2283–2296 (2014).

173. Waters, A. J. *et al.* Natural variation for gene expression responses to abiotic stress in maize. *Plant J.* **89**, 706–717 (2017).

174. Springer, N. M. & Schmitz, R. J. Exploiting induced and natural epigenetic variation for crop improvement. *Nat. Rev. Genet. 2017 189* **18**, 563–575 (2017).

175. Wang, X. *et al.* Dissecting cis-regulatory control of quantitative trait variation in a plant stem cell circuit. *Nat. Plants 2021 74* **7**, 419–427 (2021).

176. Pan, C. *et al.* CRISPR–Act3.0 for highly efficient multiplexed gene activation in plants. *Nat. Plants 2021 77* **7**, 942–953 (2021).