



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

PhD in Biotechnology

**DEVELOPMENT OF CRISPR-BASED  
PROGRAMMABLE TRANSCRIPTIONAL REGULATORS  
AND THEIR APPLICATIONS IN PLANTS**

**Sara Selma García**

Advisor: Dr. Diego Orzáez Calatayud



Valencia,  
May 2022





UNIVERSITAT POLITÈCNICA DE VALÈNCIA

PhD in Biotechnology

**Development of CRISPR-based programmable  
transcriptional regulators and their  
applications in plants**

Sara Selma García

Advisor: Dr. Diego Orzáez Calatayud

Valencia, May 2022



El Dr. Diego Orzáez Calatayud, Científico Titular del Consejo Superior de Investigaciones Científicas, perteneciente al Instituto de Biología Molecular y Celular de Plantas (IBMCP, UPV-CSIC) de Valencia, CERTIFICA que Sara Selma García, ha realizado bajo su dirección en el Instituto de Biología Molecular y Celular de Plantas, el trabajo titulado “Development of crispr-based programmable transcriptional regulators and their applications in plants”, y que autoriza su presentación para optar al grado de Doctor en Biotecnología de la Universitat Politècnica de València.

Y para que así conste, firma el presente certificado en Valencia a 11 de Mayo de 2022.



ORZAEZ  
CALATAYUD  
DIEGO  
VICENTE -  
20151834Q

Firmado  
digitalmente por  
ORZAEZ CALATAYUD  
DIEGO VICENTE -  
20151834Q  
Fecha: 2022.05.11  
17:55:31 +02'00'

Dr. Diego Orzáez Calatayud



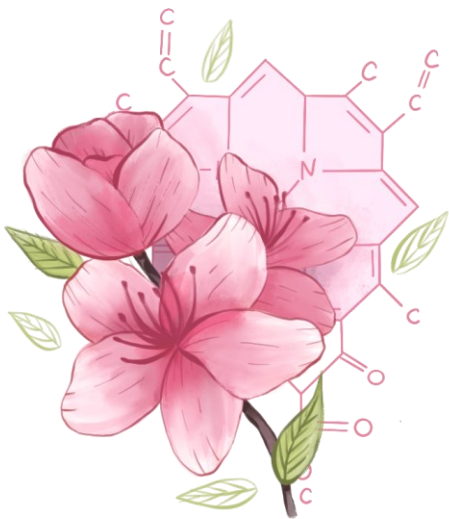
# TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS</b> .....	<b>1</b>
<b>ABSTRACT</b> .....	<b>9</b>
ABSTRACT .....	11
RESUMEN .....	13
RESUM .....	15
<b>ABBREVIATIONS</b> .....	<b>17</b>
<b>INTRODUCTION</b> .....	<b>23</b>
1    A brief introduction to Synthetic Biology .....	25
1.1    Synthetic Biology as an engineering discipline .....	26
2    The rise of plant Synthetic Biology .....	30
2.1    The standardization in cloning: modular cloning systems for plant biotech .....	32
2.2    Synthetic gene circuits in plants .....	35
2.2.1    Sensor modules for plant Synthetic Biology .....	36
2.3    Expanding the versatility of gene circuits: processors for gene regulation in plants .....	39
2.3.1    Programmable transcriptional regulators as versatile elements for the design of genetic processors .....	42
2.3.2    CRISPR tools for gene regulation .....	43
2.3.3    Transcriptional regulation with dCas in plants .....	45
<b>OBJECTIVES</b> .....	<b>47</b>
<b>CHAPTER 1</b> .....	<b>51</b>
ABSTRACT .....	53
INTRODUCTION .....	54
MATERIALS AND METHODS .....	55
RESULTS .....	60
DISCUSSION .....	78
<b>CHAPTER 2</b> .....	<b>83</b>
ABSTRACT .....	87
INTRODUCTION .....	88
MATERIALS AND METHODS .....	92
RESULTS .....	97
DISCUSSION .....	109
<b>CHAPTER 3</b> .....	<b>113</b>

ABSTRACT .....	115
INTRODUCTION .....	116
MATERIALS AND METHODS .....	119
RESULTS .....	124
DISCUSSION .....	140
<b>GENERAL DISCUSSION .....</b>	<b>145</b>
<b>CONCLUSIONS.....</b>	<b>157</b>
<b>SUPPLEMENTARY DATA.....</b>	<b>161</b>
SUPPLEMENTARY FIGURES .....	163
SUPPLEMENTARY TABLES .....	169
<b>REFERENCES.....</b>	<b>187</b>



# Acknowledgements





## ACKNOWLEDGEMENTS

*“If I'm curt with you, it's because time is a factor here. I think fast, I talk fast, and I need you guys to act fast if you want to get out of this”-Mr Wolf.*

Cuando decidí emprender este viaje, creo que no tenía muy claro donde me metía. Siempre oyes las típicas frases alarmistas como “¡Oh dios mío!, ¿estás segura de que quieres meterte ahí? Es durísimo” o “No vas a ganar para antidepresivos, piénsatelo”. Sin embargo, las de ganas de probarme a mí misma y que soy terca como una mula vieja, fue la mezcla perfecta para lanzarme de cabeza a la piscina.

No voy a decir que todos exageraban y que no ha sido tan duro como pintaban, porque estaría mintiendo vilmente. Sin embargo, pese a pasar momentos difíciles y frustrantes estoy segura de que, si me dirán a elegir, escogería el mismo camino sin dudarlo.

El labo 2.10/2.09 me ha visto crecer, me ha permitido aprender y mejorar cada día. Aquí he encontrado mi pasión y una fuerza (o cabezonería) que no conocía y que me ha impulsado a superar cada uno de los obstáculos que encontraba. Aquí he conocido gente maravillosa, que me ha acompañado durante toda esta etapa y que gracias a ellos puedo decir que este espacio entre batas y benthamianas ha sido un pequeño hogar.

Por todo esto me gustaría agradecer a todos aquellos que han estado a mi lado, de una forma o de otra y que me han ayudado a completar este viaje:

A mi director Diego Orzáez (o Sr. Jefe), gracias por darme esta oportunidad, por tu paciencia todos estos años y por guiarme en cada uno de los proyectos que emprendía. Recuerdo el primer día que llegué, completamente nerviosa y bastante perdida. Me recibiste con una sonrisa y me diste la bienvenida... y luego mucha faena. Reconozco que no soy la misma persona que entró aquel día, y supongo que en parte también es gracias a ti. Contigo he aprendido muchísimo (a día de hoy aún sigo aprendiendo) y, aunque hemos pasado momentos complicados, acabo esta etapa con un gran sentimiento de gratitud y de cariño. Gracias por tu tiempo, por explorar conmigo los

## Acknowledgements

---

limites de tu paciencia, por animarme cuando me venía abajo, por enseñarme Izal y, sobre todo, por creer en mí.

A todos mis compañeros del 2.10/2.09. Esa familia de bata blanca con la que tanto he compartido estos años: Alegrías, frustraciones, celebraciones, chaladurias, excursiones, debates de ciencia y ¡muchas cervezas! He pasado momentos estupendos con todos vosotros y os tengo mucho que agradecer:

A Silvia, mi italiana favorita, la luz de mi vida. He sido muy afortunada al a mi lado, teniendo la oportunidad conocerte mejor y encontrar a una persona estupenda. Me han encantado nuestras charlas de ciencia, nuestros viajes en coche hacia Madrid, nuestras excursiones por los almendros en flor, enseñarte Soria, descubrir nuevos restaurantes y nuestros intercambios de comidas. No solo has sido una compañera de labo, has sido una amiga durante todo este tiempo, demostrándome en los peores días que todo se podía arreglar con un “Va ¿salimos de aquí y vamos a por una cerveza?”. Un cuore freddo non è mai stato così caldo. A Asier, por estar siempre ahí, por los buenos momentos, por ser un gran amigo con el que siempre puedo contar, por aguantar mis histerismos con tu particular sentido del humor y por ayudarme con mis crisis de identidad. A Lolo, la chalada de mis amores. Aunque me hayas pillado en las etapas finales de este periodo te has convertido en alguien imprescindible, llenando los días tediosos y frustrantes de color, locura y risas (y alguna vez de tu música rara de modernos). Sigue siempre siendo tú. A Bea, por preocuparte siempre de que estuviera bien, por sonreír, aunque estuvieras agotada, por hacer del labo una piña donde todos nos sintiéramos a gusto. Eres una persona estupenda, que vale muchísimo y estoy muy agradecida de poder haberte podido conocer mejor durante estos últimos años. A Asun, por enseñarme tanto. Siempre has estado para mi cuando se me acababan las ideas o no sabía que más hacer. Me ayudaste desde que puse el pie en el laboratorio, siendo tu patito, y hasta el día de hoy me sigues enseñando. Me has conseguido tranquilizar cuando era un matojo de nervios, con tu forma práctica de ver las cosas y que tanto me ha ayudado a sacar mis proyectos adelante. A Marta Vázquez, por tener siempre un momento para ayudarme, aunque fueras hasta arriba de faena. Gracias por tus consejos, por los días de sushi, por Oporto, por las sesiones de deporte del 2.10 de las ocho y por aguantarme todos estos años sentada a tu lado diciendo improperios en mi

leguaje extraño. A Elena, por ser esa persona con la que siempre puedes contar. Gracias por tener el corazón tan grande y dar siempre lo mejor de ti para los demás. A Silvia Presa, porque, aunque todos te acabamos desquiciando, siempre nos ayudas y sin ti todo sería un caos absoluto. Eres una bellísima persona con la que siempre me he sentido cuidada y arropada durante todos estos años. A Marta (Chiquita), porque durante esta última etapa de tesis tenerte a mi lado me ha ayudado a pasar mucho mejor los malos momentos. Me ha encantado compartir piso contigo porque me ha dado la oportunidad de conocerte mejor y te he cogido muchísimo cariño. A Maria, por tantos buenos momentos y por no huir del labo por mis locuras. Víctor, compañero de frikadas, por ilustrarme con la bioinformática. Nuccio, por hacerme siempre reír cuando podía estar de bajón. Rubén, por estar siempre ahí cuando iba quejándome de algo. Jordi, por tus buenos consejos en materia de “networking” y Camilo, por las largas charlas hablado de mil cosas. Gracias por ser los mejores compañeros de labo que una pueda desear.

Por otro lado, también me gustaría agradecer a Antonio Granell por acogerme todo este tiempo en su laboratorio y por sus buenos consejos, a todo el equipo de Jose Antonio Darós, especialmente a Mireia, por las colaboraciones y todo lo aprendido con los sistemas virales, y al equipo de metabolómica, por todo lo que me han enseñado y la paciencia que han tenido.

Also, I would like to thank Mathew Lewsey and his team for welcoming me into their lab and making me feel like one of the team. I will always remember my stage in Melbourne with great affection.

Más allá del ámbito del laboratorio hay mucha gente que me ha apoyado y que me gustaría agradecer, porque sin ellos todo esto no habría sido posible. Primero, a mis padres. Pare y mamá, esta tesis es completamente para vosotros. Cada letra, cada experimento, cada publicación, cada logro os lo debo a vosotros. Habéis creído en mi cuando yo ya no podía más, me habéis dado ánimos cuando estaba más hundida y me habéis hecho reflotar, me habéis apoyado en cada paso que daba, desde que era una niña hasta que crecí y os dije que quería hacer cosas raras en un laboratorio. Hebu fet de mi tot el que sóc i no podria estar més feliç de compartir aquest somni amb vosaltres.

A mi tía Nati, por ser ese pilar tan importante de mi vida. Desde mis primeros pasos siempre me cuidaste, enseñaste y protegiste. Cuando crecí me seguiste cuidando todo el camino, preocupándote por mí y estando siempre ahí, dándome siempre palabras de ánimo y la más dulce de las sonrisas que me dieron fuerzas para enfrentarme a lo que fuera. A mi tía Cristi, gracias por los buenos consejos, por las largas charlas, por apoyarme siempre, por evitar que muriera de inanición durante el último periodo de tesis, por todas las historias que me contabas de niña que me hicieron querer saber y conocer más del mundo que me rodeaba y por todo el cariño que siempre me has dado. A ma tia Vivi, gràcies per haver-te preocupat per mi tant tots aquests anys. Per acceptar-me amb les meves bogeries i animar-me sempre. A mon tio Felipe, gràcies per germinar en mi el gust per la música i per aquests concerts que em feien allunyar-me de l'estrès i la frustració del dia a dia. A Terebela, ets el més semblant a una germana gran que tinc. gràcies per donar-me suport sempre.

A Neus i Carles, sou els millors amics que una es pot imaginar. Sense vosaltres, literalment, aquesta tesi no hauria tirat endavant. Gràcies per cuidar-me tant, per entestar-vos que jo continués estable encara en els pitjors moments. Gràcies per transformar-vos en una família per a mi i fer-me somriure quan tot era fosc. A Alberto, per ser el meu gegant de cor d'or. Graciés per aguantar els meus plors i frustracions amb la teva calidesa habitual. A July, porque has estado a mi lado durante todo este periodo, ayudándome a pasar los peores momentos y a centrarme cuando entraba en mis bucles de frustración. A Olga y Quico, por todos los buenos momentos y por los viajes que hicieron que toda etapa fuera mucho más fácil. Gracias por preocuparos y estar siempre ahí para mí. A mis sorianos, Esther, Rodri y Miky, por transportarme lejos y hacer que todo el tiempo que estoy con vosotros esté en paz y tranquila. Gracias por estar siempre ahí, por muchos años que pasen o por muchos kilómetros que nos separen. A Vicente, por hacerme reír siempre, por ser mi torpe chalado favorito con el que tantos buenos ratos he pasado. A Iván, porque ¿quién no quiere a Iván? Gracias por esas largas charlas de ciencia y compartir las penas de la tesis.

Finalmente, quiero agradecer a Andrés, por ser el mejor compañero de viaje. No sé qué hubiera hecho sin ti, sobre todo en este final de tesis. Has sido esa persona con la que siempre he podido ser yo. Has visto todas las Saras de mi cabeza, todas las chaladuras

que normalmente mantengo atadas y me has aceptado siempre con cariño y ... también con algo de locura. Gracias por hacerme sonreír día tras día y creyendo en mí cuando ni siquiera yo lo hacía.

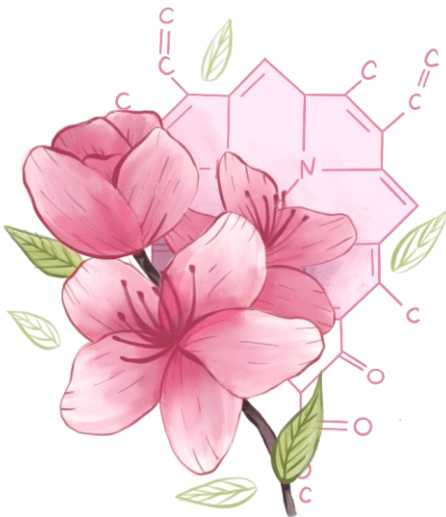
Este capítulo llega a su fin y aunque ahora mismo siento una sensación agri dulce en mi interior, estoy segura de que esto no es una despedida. No solo me voy con todo lo aprendido, si no con la certeza de que he conseguido ser una mejor versión de mí misma. Lo mejor de todo es que mi siguiente capítulo empezará cargado de recuerdos y de lazos con gente maravillosa que atesoraré toda la vida. A todos vosotros, recordad siempre con todo mi cariño: *Me agotáis.*







# Abstract





## ABSTRACT

Plant Synthetic Biology aims to redesign plants to acquire novel traits and functionalities through orthogonal regulatory circuits. To achieve this goal, new molecular tools with the capacity of interacting with endogenous factors in a potent and specific manner must be developed. CRISPR/Cas9 emerged as promising tools which combine a customizable DNA-binding activity through the catalytically inactivated version of Cas9 protein (dCas9) with the possibility to anchor autonomous transcriptional activation domains (TADs) to its structure to achieve a specific regulation of the gene expression. The Programmable Transcriptional Activators (PTAs) could act as specific, orthogonal and versatile processor components in the development of new genetic circuits in plants. In search for optimized dCas9-PTAs, a combinatorial evaluation of different dCas9 architectures with a catalogue of various TADs was performed. The best resulting tool of this comparison, named dCasEV2.1, is based on the scRNA strategy and the combination of EDLL and VPR activation domains with a multiplexable gRNA2.1 loop, which is a mutated version of the previously described gRNA2.0. In this work, the dCasEV2.1 programmable activator was proved to be a strong and specific tool, achieving higher activation rates than other available dCas9 strategies in plants. Unprecedented activation rates were observed targeting endogenous genes in *N. benthamiana*, accompanied by strict genome-wide specificity that makes this tool suitable to perform a tight regulation of complex regulatory networks. As a proof of concept, a design of four activation programs to activate different branches of the flavonoid pathway and obtain specific metabolic enrichments in *N. benthamiana* leaves was performed. The metabolic analysis on the dCasEV2.1 metabolically reprogrammed leaves revealed a selective enrichment of the targeted metabolites and their glycosylated derivatives that correlated with the activation program employed. These results demonstrate that dCasEV2.1 is a powerful tool for metabolic engineering and a key component in genetic circuits aimed at reprogramming metabolic fluxes. Finally, based on dCasEV2.1, we developed an optimized Viral Induced Gene Regulation (VIGR) system that makes use of a Potato Virus X (PVX) vector for the delivery of the gRNA-encoded CRISPR activation

programs. This approach offers a way to control the plant transcriptome through a spray-based systemic delivery of CRISPR components to adult plants. The new PVX-VIGR system led to strong transcriptional activation in several endogenous target genes, including three selected MYB-like transcription factors. Specific MYB activations lead to distinctive metabolic profiles, showing that the potential applications of the dCasEV2.1 tool in plants include the obtention of custom metabolic profiles using a spray-based delivery of gRNA-encoded transcriptional reprogramming instructions. In sum, this thesis provides novel tools for strong, orthogonal and programmable transcriptional activation in plants, with an expanded toolbox for the delivery of the activation programs.

## RESUMEN

La Biología Sintética de Plantas tiene como objetivo rediseñar las plantas para que adquieran características y funcionalidades novedosas a través de circuitos reguladores ortogonales. Para lograr este objetivo, se deben desarrollar nuevas herramientas moleculares con la capacidad de interactuar con factores endógenos de manera potente y específica. CRISPR/Cas9 surgió como una herramienta prometedora que combina la capacidad personalizable de unión al DNA, a través de la versión catalíticamente inactivada de la proteína Cas9 (dCas9), con la posibilidad de anclar dominios autónomos de activación transcripcional (TADs) a su estructura para lograr una regulación específica de la expresión génica. Los activadores transcripcionales programables (PTAs) pueden actuar como procesadores específicos, ortogonales y versátiles para el desarrollo de nuevos circuitos genéticos en las plantas. En busca de dCas9-PTA optimizados, se llevó a cabo una evaluación combinatoria de diferentes arquitecturas dCas9 con un catálogo de varios TAD. La mejor herramienta resultante de esta comparación, denominada dCasEV2.1, se basa en la estrategia scRNA y la combinación de los dominios de activación EDLL y VPR con un bucle multiplexable gRNA2.1, que es una versión mutada del gRNA2.0 descrito previamente. En este trabajo, el activador programable dCasEV2.1 demostró ser una herramienta potente y específica, logrando tasas de activación más altas que otras estrategias dCas9 disponibles en plantas. Se observaron tasas de activación sin precedentes dirigidas a genes endógenos en *N. benthamiana*, acompañadas de una estricta especificidad en todo el genoma, lo que hace que esta herramienta sea adecuada para la regulación estricta de redes reguladoras complejas. Como prueba de concepto, se diseñaron cuatro programas de activación para distintas ramas de la ruta de los flavonoides, buscando obtener enriquecimientos metabólicos específicos en hojas de *N. benthamiana*. El análisis metabólico de las hojas metabólicamente reprogramadas mediante dCasEV2.1 reveló un enriquecimiento selectivo de los metabolitos diana y sus derivados glicosilados, que se correlacionaron con el programa de activación empleado. Estos resultados demuestran que dCasEV2.1 es una herramienta eficaz para la ingeniería metabólica y un componente clave en los circuitos genéticos destinados a reprogramar los flujos metabólicos. Finalmente,

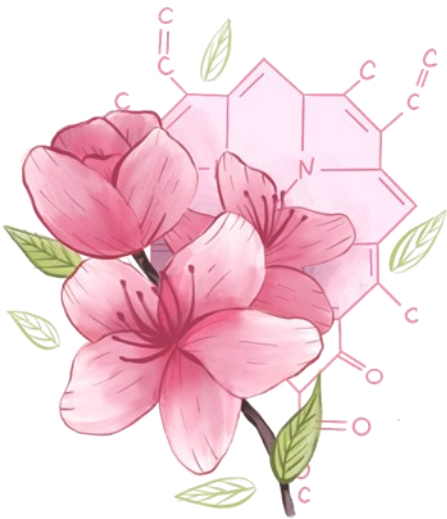
basándonos en dCasEV2.1, desarrollamos un sistema optimizado de regulación de genes inducidos por virus (VIGR) que utiliza un vector Potato Virus X (PVX) para el suministro de los programas de activación CRISPR codificados con gRNA. Este enfoque permite controlar el transcriptoma de la planta a través de una aplicación sistémica basada en aerosol de componentes CRISPR a plantas adultas. El nuevo sistema PVX-VIGR produjo una fuerte activación transcripcional en varios genes diana endógenos, incluidos tres factores de transcripción MYB-like seleccionados. Las activaciones específicas de MYB condujeron a perfiles metabólicos distintivos, demostrando que las aplicaciones potenciales de la herramienta dCasEV2.1 en plantas incluyen la obtención de perfiles metabólicos personalizados utilizando un suministro basado en aerosol de instrucciones de reprogramación transcripcional codificadas por gRNA. En resumen, esta tesis proporciona herramientas novedosas para la activación transcripcional fuerte, ortogonal y programable en plantas, con una caja de herramientas ampliada para el suministro de los programas de activación.

## RESUM

La Biologia Sintètica de Plantes té com objectiu redissenyar les plantes per que obtinguen característiques i funcionalitats innovadores mitjançant circuits reguladors ortogonals. Per arribar a aquest objectiu, s'han de desenvolupar noves ferramentes moleculars amb la capacitat d'interactuar amb factor endògens d'una manera potent i específica. CRISPR/Cas9 va sorgir com una ferramenta prometedora que combina la capacitat personalitzable d'unió al DNA, mitjançant la versió catalíticament inactivada de la proteïna Cas9 (dCas9), amb la possibilitat de fixar dominis autònoms de activació transcripcional (TADs) a la seua estructura per aconseguir una regulació específica de la expressió gènica. Els activadors transcripcionals programables (PTAs) poden actuar com a processadors específics, ortogonals i versàtils per al desenvolupament de nous circuits genètics a les plantes. Buscant dCas9-PTA optimitzats, es va realitzar una avaluació combinatòria de distintes arquitectures dCas9 amb un catàleg de diversos TAD. La millor ferramenta segons aquesta comparació, anomenada dCasEV2.1, es basa en la estratègia scRNA i la combinació del dominis d'activació EDLL i VPR amb un bucle multiplexable gRNA2.1, que es una versió mutada del gRNA2.0 descrit prèviament. En aquest treball, el activador programable dCasEV2.1 es va mostrar com una ferramenta potent i específica, aconseguint nivells d'activació majors que altes estratègies dCas9 disponibles en plantes. Es van observar taxes d'activació sense precedents dirigides a gens endògens en *N. benthamiana*, junt a una estricta especificitat en tot el genoma, indicant que aquesta ferramenta és adequada per a la regulació estricta de xarxes reguladores complexes. Como prova de concepte, se van dissenyar quatre programes d'activació per a diferent branques de la ruta dels flavonoides, cercant obtenir enriquiments metabòlics específics en fulles de *N. benthamiana*. L'anàlisi metabòlic de les fulles metabòlicament reprogramades mitjançant dCasEV2.1 va revelar un enriquiment selectiu del metabòlits diana i els seus derivats glicosilats que es correlacionen amb el programa d'activació emprat. Aquests resultats demostren que dCasEV2.1 és una ferramenta eficaç per a l'enginyeria metabòlica i un component clau als circuits genètics destinats a reprogramar els fluxos metabòlics. Finalment, en base a dCasEV2.1, desenvoluparem un sistema optimitzat de regulació de gens induïts per virus (VIGR) que utilitza un vector

Potato Virus X (PVX) per al subministrament dels programes d'activació CRISPR codificats amb gRNA. Aquesta aproximació permet controlar el transcriptoma de la planta mitjançant l'aplicació sistèmica basada en aerosol de components CRISPR a plantes adultes. El nou sistema PVX-VIGR va produir una gran activació transcripcional en diversos gens diana endògens, inclosos tres factors de transcripció MYB-like seleccionats prèviament. Les activacions específiques de MYB conduïren a perfils metabòlics distintius, demostrant que les aplicacions potencials de la ferramenta dCasEV2.1 en plantes inclouen la obtenció de perfils metabòlics personalitzats emprant un subministrament basat en aerosol de instruccions de reprogramació transcripcional codificades per gRNA. En resum, aquesta tesi proporciona noves ferramentes per a l'activació transcripcional forta, ortogonal i programable en plantes, amb una caixa de ferramentes eixamplada per al subministrament dels programes d'activació.





# Abbreviations



## ABBREVIATIONS

4CL: 4-coumaroyl CoA ligase

ABA: abscisic acid

AD: action domains

aTc: anhydrotetracycline

bp: base pair

Bt: *Bacillus thuringiensis*

C4H: Cinnamate 4-hydroxylase

Cas9-PTAs: CRISPR-dCas9 Programmable Transcriptional Activators

cDNA: complementary DNA

CHI: Chalcone isomerase

CHS: Chalcone synthase

CP: coat proteins

crRNA: CRISPR RNA

dCas9: dead Cas9

dCasEV2.1: dCas9:EDLL-MS2:VPR/gRNA2.1

DDA: data-dependent acquisition

DFR: Dihydroflavonol 4-reductase

DNA: deoxyribonucleic acid

dpi: days post infiltration

ERF: Ethylene Response Factor

## Introduction

---

ESI: electrospray ionization

F3'H: *Flavonoid 3'-hydroxylase*

F3H: Flavanone 3-hydroxylase

FDR: false discovery rate

FLS: Flavonol synthase

Fluc: firefly luciferase

GB: GoldenBraid

GFP: green fluorescent protein

GMO: genetically modified organisms

gRNA: guide of RNA

h: hours

IGEM: International Genetically Engineered Machines

IPTG: isopropyl- $\beta$ -D-thiogalactopyranoside

JA: jasmonic acid

LI: local infection

Luc: luciferase

MES: 2-(N-morpholino)ethanesulfonic acid

min: minutes

MS: Murashige and Skoog medium

Nos: Nopaline synthase

OD: Optical density

PAL: Phenylalanine ammonia-lyase

PAM: protospacer adjacent motif

PBP: periplasmic binding protein

PCA: Principal Component Analysis

PCR: polymerase chain reaction

pNos: Nopaline synthase promoter

pre-crRNA: precursor crRNA

PTA: Programmable Transcriptional Activator

PTRs: programmable transcriptional regulators

PVX: Potato virus X

Rluc: renilla luciferase

RNA: ribonucleic acid

RNPs: ribonucleoproteins

RPUs: reference Promoter Units

RTAs: relative transcriptional activities

SA: salicylic acid

SAM: synergistic activation mediator

SBOL: Synthetic Biology Open Language

scRNAs: CRISPR scaffold RNAs

SD: standard deviation

sgRNA: small RNA guide

SI: systemic infection

SID: mSin3 interaction domain

## Introduction

---

SRA: Sequence Read Archive

STD: standard

SynBio: Synthetic Biology

SYNV: Sonchus yellow net rhabdovirus

TAD: transcriptional activator domains

TALE: transcription activator-like effector

T-DNA: Ti-Plasmid from Agrobacterium

TetR: tetracycline repressor

TF: transcriptional factors

TMV: Tobacco mosaic virus

tracrRNA: trans-activating

tRNA: transfer RNA

TRV: Tobacco rattle virus

TSS: Transcriptional Start Site

TUs: Transcriptional Units

VIGA: virus-induced gene activation

VIGE: virus-induced gene editing

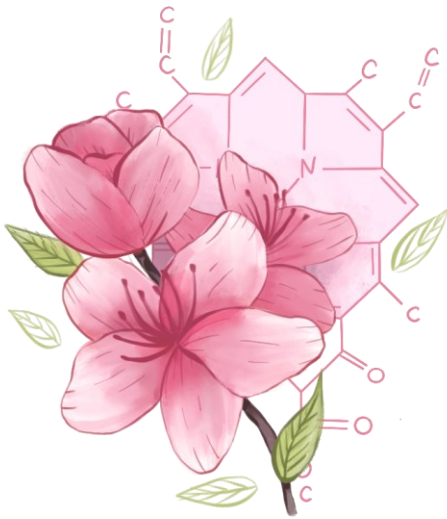
VIGR: Virus-Induced Gene Reprogramming

VIGS: virus induced-gene silencing

VPR: VP64-P65-RTA

ZF: zinc finger

ZFN: zinc finger nuclease



# Introduction





# INTRODUCTION

## 1 A brief introduction to Synthetic Biology

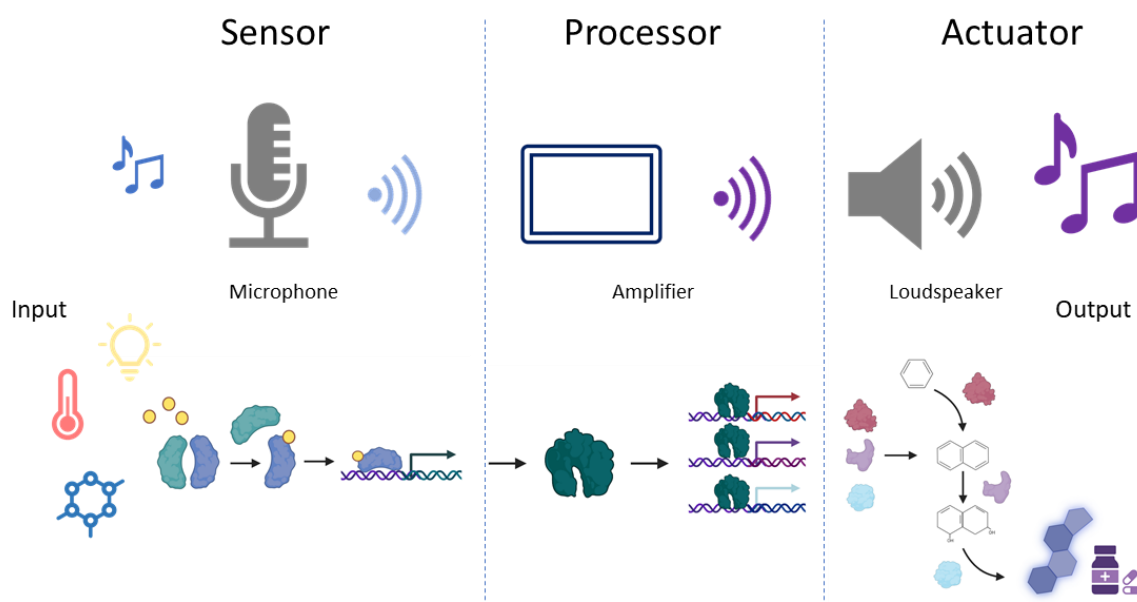
Synthetic Biology (SynBio) is a fast-evolving discipline that combines engineering with biology to obtain new biological systems with useful characteristics and applications. This field, which emerged as a response to the current society's needs in biomedicine, agriculture and manufacturing, applies the engineering principles of standardization, modularity and abstraction to the design of new and/or modified biological systems. The so-called bottom-up approaches in SynBio aim at creating new biological systems from very basic chemical components, redesigning the biological system from its foundations (Smith *et al.*, 2003; Koide *et al.*, 2009). A less radical group of approaches, often known as top-down approaches, aim at employing current cellular organisms as “chassis”, on top of which new functions are engineered. The objective of the top-down SynBio strategies is to build inside the cell new genetic “devices” made of gene circuits capable of generating predictable and efficient outputs, through user-designed inputs (Benner and Sismour, 2005; Khalil and Collins, 2010). Typically, the creation of the new biological devices proceeds through cycles of improvement that encompass four characteristic steps: design, build, test and learn (DBTL) (Slusarczyk *et al.*, 2012; Way *et al.*, 2014; Carbonell *et al.*, 2018). The design step comprises the conceptualization of the new genetic device, establishing the desired inputs and outputs to select optimized network hierarchies, kinetic parameters, and genetic parts. The design step includes a modelling process to predict the circuit behaviour *in silico*, its sensitivity and robustness. The build step encompasses the standardized cloning of the selected genetic parts and their integration into the biological system (chassis). The test step comprises the *in vivo* evaluation of the genetic circuit and the optimisation of its employment. Finally, the learn step includes the downstream analysis of the results obtained, employing statistical methods and predictive models, and using them to optimize future designs in subsequent DBTL cycles.

## 1.1 Synthetic Biology as an engineering discipline

The roots of synthetic biology in its top-down version date back to the emergence of recombinant DNA technology. The first genetically modified microorganisms (Hobom, 1980) provided for the first time control over the functions of a living organism through the construction of recombinant genetic elements (Cameron *et al.*, 2014). However, the concept of synthetic biology was not framed in the scientific context until the creation of synthetic genetic circuits and regulatory networks that allowed the cell to respond to the environment predictably. This was possible thanks to the convergence and interconnection of the advances in genome sequencing techniques, DNA synthesis, and computational tools, giving scientists a global vision of the mechanism of cellular machinery (Purnick and Weiss, 2009). Employing a rational design that was reminiscent of the engineering context, basic genetic parts, such as promoters, coding sequences or terminators, which are the basic constituents of the architecture of the genetic circuits, started to be characterized and documented by early synthetic biologists in an analogous way as electronic parts are documented in electronic devices (Haseloff and Ajioka, 2009). In one more step of complexity, these genetic elements were combined for the construction of functional genetic modules, catalogued as sensors, processors and actuators, which represented a biological function with the ability to act independently of the biological context (De Las Heras *et al.*, 2010; Qi *et al.*, 2013a). The combination of these functional modules in an organism (chassis) expands its ability to process information and integrate new types of stimuli to provide an expected response or output.

Maintaining the analogy, an electronic system comprising the above-mentioned sensor, processor and actuator set could be, for example, a sound amplification device. The microphone, which would act as a sensor, detects the sound signal that is sent to the amplifier. The amplifier, acting as a processor, transforms the signal, increasing the potency of the signal, which is sent to the loudspeaker, an actuator system that generates a sound output with the desired characteristics. Similarly, in a biological system, the sensor inputs non-transcriptional signals from e.g., the environment, such as light intensity, temperature changes or the presence of a certain chemical compound

and generates a transcriptional signal as output. A processor inputs that transcriptional signal(s) from a sensor(s) and transforms it into a different transcriptional output. Finally, the actuator inputs transcriptional signals from a processor and generates a non-transcriptional response (Rollié *et al.*, 2012; Figure 1).



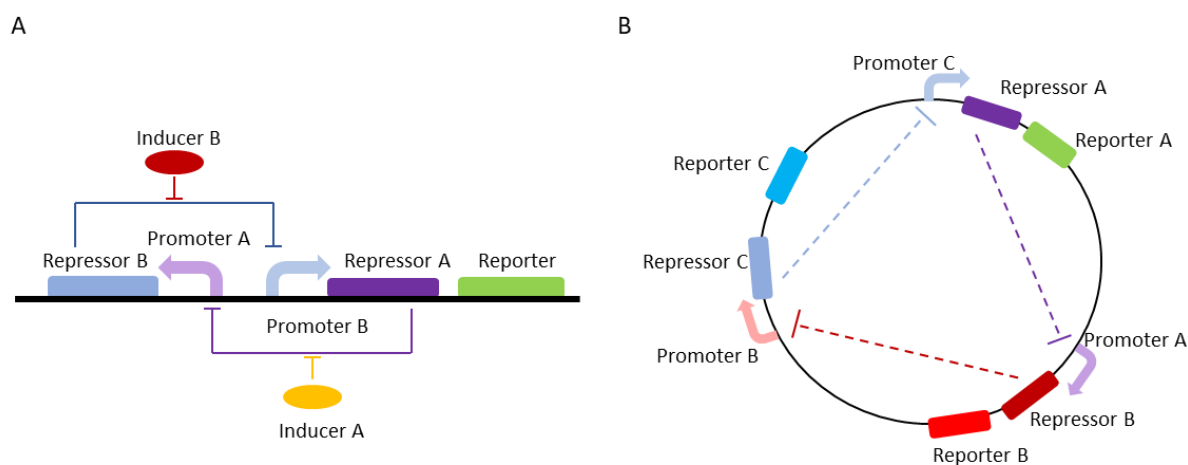
**Figure 1: Analogy of the sensor-processor-actuator scheme in biological systems.** The sensor, processor and actuator modules are represented by an electronic and a genetic device. In electronics, sensors receive non-electronic inputs and transform them into electronic signals that are sent to processors. These transform the input electronic signal into an output electronic signal with different characteristics, which is then transferred to the actuator. Finally, actuators generate a non-electronic response or output. In biological circuits, sensors receive non-transcriptional inputs such as light, temperature, or the presence of a chemical inducer, and turn them into transcriptional signals. Processors transform transcriptional inputs into new modified (e.g., integrated or amplified) transcriptional outputs, and finally, actuators receive the new transcriptional input and convert it into a non-transcriptional output, such as the activation of a metabolic pathway that brings to the synthesis of a new chemical.

The first examples of synthetic genetic circuits that framed this engineered-based methodology were the genetic toggle switch (Gardner *et al.*, 2000) and the repressilator (Elowitz and Leibler, 2000), both elaborated in *Escherichia coli*. The genetic toggle switch is a bistable synthetic gene circuit that consists of two genes that mutually repress each other. Two possible equilibrium states in the cell are possible since the expression of

one of the constituent genes is inhibited by the transcription product of the opposing gene. Consequently, the toggle switch presents a strong “memory” of the equilibrium state, since it does not flip randomly between states. Once equilibrium has been reached, it can only be changed to the alternate state by the addition of a specific ligand that inhibits the action of the repressor protein (Figure 2A). The first genetic toggle switch incorporates the mutually repressed genes, *lacI* and *tetR*, coupled to a fluorescent reporter system. By the addition of the specific ligands, isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) and anhydrotetracycline (aTc) respectively, the repression generated over the promoter is inhibited, allowing the expression of the opposite gene (Figure 2A), which itself is a repressor of the first gene. In the toggle circuit, the repressor elements *LacI* and *TetR* act as sensors, taking a non-transcriptional input (ligand) and generating a (negative) transcriptional response. The pair of promoters integrating the two mutually repressing signals jointly make up the processor of the circuit. The integrated transcriptional output produced by the mutually repressed transcriptional units is then taken by the reporter gene that acts as a model actuator. Some examples of applications of the genetic toggle switch are the development of glucose memory sensors (Bothfeld *et al.*, 2017) or their use for the production and delivery of therapeutic molecules (Din *et al.*, 2016; Wegmann *et al.*, 2017).

The repressilator is a gene circuit that operates in cyclic feedback loops of repression, in an equivalent way to how the toggle switch works. The objective behind this design is to obtain a genetic circuit that generates an oscillation state of its internal components. Although it is possible to operate with a single element that represses itself and generates a stable fluctuation in its expression (Dilão, 2014), the first repressilator described by Elowitz and Leibler (2000) was composed of three repressive elements, *LacI*, *TetR* and *Cl* genes, operating cyclically. The first repressor (*LacI*) represses the transcription of the next repressor gene (*TetR*), whose expression product represses the expression of the third gene (*Cl*). To close the cycle, the *Cl* expression product inhibits the *LacI* (Figure 2B). In this case, sensor (repressor protein) levels are controlled by the transcriptional output associated with its linked promoter, whose input is determined by the action of the sensor of the previous repressive element. The expression of the different repressors of the circuit can also be coupled to a reporter system, which

operates as an actuator to monitor the oscillations produced. The repressilator design was extracted from the biological oscillators found in nature, such as the circadian clock (Bell-Pedersen *et al.*, 2005) or the cell cycles (Pomerening *et al.*, 2005), and could be employed for modelling complex regulatory networks and studying how different components interact in cells (Chen and Aihara, 2002; Garcia-Ojalvo *et al.*, 2004; Strelkova and Barahona, 2011).



**Figure 2: Representation of the first genetic circuits in *E. coli*.** A) Schematic representation of the genetic toggle switch, which is a bistable circuit with memory that comprises two genes that mutually repress each other. B) Schematic representation of the repressilator, which is a circuit that generates an oscillation state of their components through the alternate expression of proteins that cyclically repress each other.

These genetic circuits represented the starting point for the design of increasingly sophisticated synthetic devices, such as other genetic switches (Deans *et al.*, 2007; Rhodius *et al.*, 2013), logic gates (Wang *et al.*, 2011; Goñi-Moreno and Amos, 2012), or counters (Friedland *et al.*, 2009). Thanks to the implementation of new sensors, processors and actuators, the range of inputs and outputs that biological systems are capable of processing are notably expanded. In addition, although the bases of synthetic biology were conceived for bacterial systems, their use is currently expanding to additional biological systems, as is the case of yeast (Liu *et al.*, 2019), mammalian cells (Shakiba *et al.*, 2021) and photosynthetic organisms (Vavitsas *et al.*, 2019). The use of photosynthetic organisms as optimized "chassis" expands the range of applications in various fields such as biosensing (Hicks *et al.*, 2020), production of pharmaceuticals (Breitling and Takano, 2015), biofuels (Georgianna and Mayfield, 2012), new biomaterials (Keating and Young, 2019), and drug discovery (Klein *et al.*, 2014).

## 2 The rise of plant Synthetic Biology

The plant kingdom represents one of the most valuable resources of food, materials and bioactive compounds, and therefore crop improvement for increased agricultural productivity has always been a challenge for plant researchers (Gosal *et al.*, 2010; Davies, 2010; Fang *et al.*, 2019). The expansion of SynBio beyond microbial systems towards alternative biological chassis has put plants in the spotlight as attractive hosts for synthetic genetic circuits and metabolic engineering approaches (Pouvreau *et al.*, 2018). The introduction of synthetic biology principles in plant biotechnology could facilitate the implementation of the complex genetic designs required to reach new challenging goals in crop productivity (Zurbriggen *et al.*, 2012; Liu and Stewart, 2015).

The employment of recombinant DNA techniques and heterologous gene expression to obtain genetically modified crops represented the beginnings of plant "customization" for generating new predictable biological functions and obtaining new agricultural and biotechnologically interesting traits. Bt crops (Estruch *et al.*, 1997; Shelton *et al.*, 2013) are genetically modified to express the toxin of *Bacillus thuringiensis* (Bt), or Roundup Ready crops resistant to the herbicide glyphosate (Shah *et al.*, 1986; Dill, 2005), are

pioneering examples of relatively simple genetic engineering approaches that have had a broad impact on crop yield worldwide, although not without rising and equally broad political controversy around genetically modified organisms (GMO).

However, the challenge of using plants as chassis for synthetic biology-like genetic engineering required the development of new methodologies, more efficient than those employed in traditional plant GMO technology. The development of sophisticated genetic circuits in plants needs to be based on the same cyclic DBTL pipeline postulated for the microbial systems (Liu and Stewart, 2015), which requires a systematic process of identifying and abstracting genetic elements refined to be orthogonal and context-independent (Schaumberg *et al.*, 2016). Even before this, it was necessary to develop new molecular tools and cloning techniques adapted to plants that enable the assembly of complex DNA constructs in a standardized manner. This process started with pioneering efforts of identifying, cataloguing and adapting for engineering uses new basic genetic parts, such as promoters, coding regions, terminators, etc. beyond those widely used, like the CaMV35S promoter and the Nopaline synthase (Nos) terminator (Lucks *et al.*, 2008; McCarthy and Medford, 2020; Amack and Antunes, 2020). The orthogonality of these basic genetic pieces is an essential characteristic for using them in a great variety of plant chassis (Kassaw *et al.*, 2018), the list of which has increased remarkably thanks to the growing collection of sequenced plant genomes (Stewart *et al.*, 2018). In addition to a large battery of new promoters and terminators driving the expression of selected genes, either constitutively or in a tissue-specific manner (Brückner *et al.*, 2015; Rudge *et al.*, 2016), basic genetic elements from other biological systems have also been identified and adapted to plants. Widely used examples of this are the fluorescent proteins employed as reporters, which facilitate the characterization of other genetic parts and the modelling of genetic circuits (Haseloff and Siemerling, 2006).

## 2.1 The standardization in cloning: modular cloning systems for plant biotech

A key initial step towards the implementation of SynBio engineering principles in plant biotech consisted of the adaptation of DNA cloning systems to the modularity and standardization principles. One of the first standardization attempts was carried out within an iGEM (international Genetically Engineered Machines) project (Smolke, 2009). In this pioneering project, plant DNA functional elements were adapted for the first time to the BioBricks cloning standard, the first DNA cloning method which established standard rules for the assembly of genetic parts (Boyle *et al.*, 2012). Despite its elegance, the BioBricks cloning system presented some limitations, such as the difficulty to generate large DNA constructs and the unavoidable presence of short “scar” sequences between the assembled parts due to the type of restriction enzymes used (type II restriction enzymes). The BioBricks cloning standard was followed by many others that contributed to facilitating the assembly of multigene DNA constructs for plant transformation such as Gibson, BglBricks or the AQUA cloning system (Gibson *et al.*, 2009; Anderson *et al.*, 2010; Beyer *et al.*, 2015). More recently, the attempts of standardization of plant genetic parts received a new push with the creation of the so-called PhytoBricks syntax (Patron *et al.*, 2015), a set of DNA cloning rules applicable to modular cloning strategies, all using type IIS restriction enzymes, such as MoClo (Weber *et al.*, 2011), GoldenBraid (Sarrion-Perdigones *et al.*, 2011) or Loop assembly (Pollak *et al.*, 2019). The adoption of the PhytoBrick standard facilitates the exchange of parts and enables a fair functional comparison of the specifications of standard plant genetic parts (PhytoBricks).

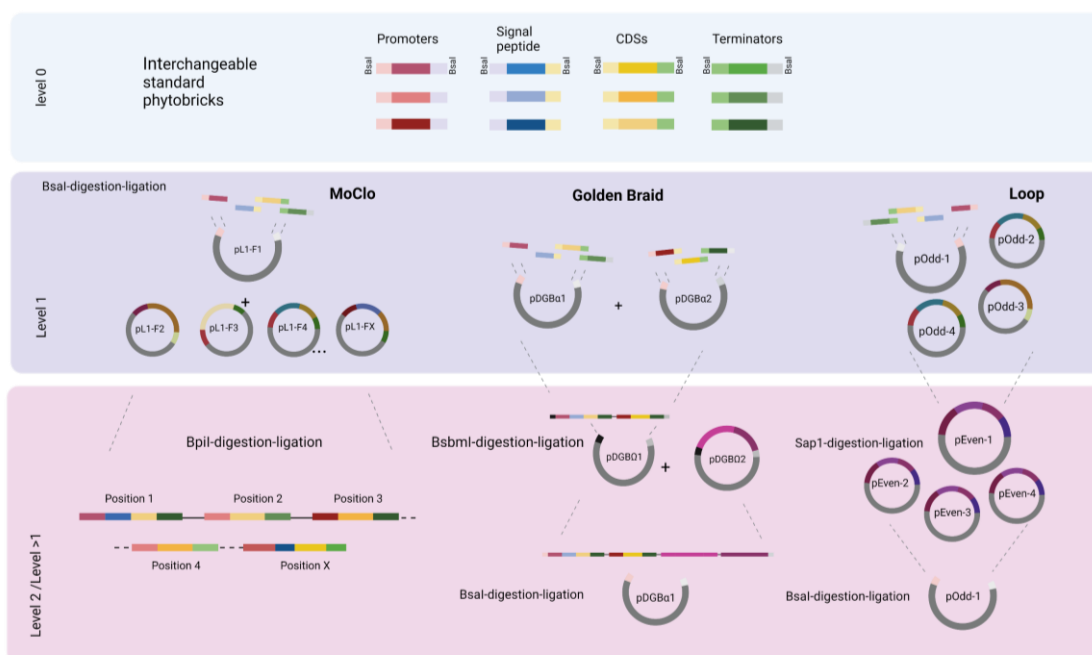
The PhytoBricks syntax operates on modular cloning systems using type IIS restriction enzymes. Modular cloning differs from traditional cloning strategies in that it builds multi genetic constructs made of several transcriptional units (TUs) using basic categorized DNA elements, which are modularly assembled employing a hierarchical set of destination vectors (Level 0, 1 and >1), and using a limited number of Type IIS restriction enzymes (usually only two). The most basic phytobrick category, involving e.g., promoters, terminators, or coding region sequences, are cloned in level 0 vectors.



These Level 0 parts are assembled forming TUs cloned in Level 1 vectors. Level 1 TUs can be assembled with other TUs to obtain complex genetic constructs (cloned in Level >1 vectors). MoClo, the first modular cloning system developed for plants, offered the possibility of assembling a pre-established number of Level 1 TUs in a single cloning step using a pre-selected destination level 2 vector (Weber *et al.*, 2011). More recently, MoClo Level 2 vectors have acquired more flexibility with the possibility of performing additional iterations of Level 1 assembly (Iverson *et al.*, 2016). The GoldenBraid (GB) cloning system, appearing soon after MoClo, offered as an alternative an iterative cloning strategy for the assembly of multigene constructs (Sarrion-Perdigones *et al.*, 2011). Although it requires more cloning steps than MoClo for the same purpose, GB offers more flexibility since all multigene constructs can be used in subsequent assemblies, allowing, in principle, endless rounds of multigene cloning. The latest updates to the GoldenBraid system (GB 4.0), implement an expanded variety of destination vectors that offer advantages in plant transformation and genome editing (Vazquez-Vilar *et al.*, 2020). The Loop assembly is the most recently developed multipartite assembly based on the same principle of IIS-type enzymes, with an iterative loop strategy consisting of two sets of four plasmid vectors (Pollak *et al.*, 2019). The lower level of assembly shares the same principle of Level 0 parts with a PhytoBricks syntax, which can be assembled into Level 1 TUs. Four Level 1 vectors are assembled into one of the four available Level 2 vectors, which can be combined to reach a Level 3 assembly comprising up to sixteen TUs. Similar to GoldenBraid, this loop process can be repeated endlessly, even interspersing vectors of different levels in the same reaction and generating virtually unlimited options of genetic combinations (Figure 3).

The development of new cloning strategies was, sometimes, accompanied by software tools for *in silico* assembly, and by the creation of databases and repositories where the phytobricks could be stored and shared with the scientific community. An example is the GoldenBraid software tool (<https://gbccloning.upv.es/>), which combines *in silico* cloning tools with a database where all phytobricks are catalogued, or the iGEM repository ([www.parts.igem.org](http://www.parts.igem.org)), a website with an integrated cloning software that includes the genetic standard parts generated in the iGEM projects, with good characterization and chassis information. Another initiative that integrates the workflow

of plant synthetic biology is the OpenPlant initiative ([www.openplant.org](http://www.openplant.org)).



**Figure 3. Representation of MoClo, GoldenBraid and Loop modular cloning systems with a hierarchical design of destination vectors.** The level 0 parts that comprise the basic genetic elements (promoters, terminators, CDS, etc.) can be interchangeable between the cloning systems. MoClo offers the possibility of assembling a pre-established number of TUs (level 1) employing the pL1-F vectors in a single cloning step using a pre-selected destination level 2 vector. GoldenBraid offers an iterative cloning strategy for the assembly of multigene constructs. The Level 1 TUs are assembled into pDGB $\alpha$ 1/2 vectors which are combined in a pDGB $\Omega$ 1/2 (Level >1). The next assemblies can be combined indefinitely using loops of pDGB $\alpha$  and pDGB $\Omega$  iterations. Loop assembly is also an iterative loop strategy consisting of two sets of four plasmid vectors (pOdd and pEven). Four Level 1 vectors are assembled into one of the four available Level 2 vectors, which can be combined to reach a Level 3 assembly comprising sixteen TUs.

This consortium had the objective of accelerating the development of new molecular tools for “engineering” plants, creating a catalogue of Phytobricks open to the scientific community, implementing automatic DNA assemblies and software for modelling gene circuits and generating plant genetic devices to meet the actual demands in agriculture, bioproduction, and bioremediation (Patron, 2016; Tuncel *et al.*, 2019; Stephenson and

Osbourn, 2020). Another interesting resource that integrates a wide range of assembly systems, as well as genetic circuit design and modelling systems, is SBOL (The Synthetic Biology Open Language). Although this open language initiative is not focused on plants, it has several advantageous features, such as the visualization of circuit design using standard glyphs ([www.sbolstandard.org](http://www.sbolstandard.org)) and the direct sequence submission to synthetic biology repositories such as ICE ([www.jbei.org](http://www.jbei.org)) or SynBio hub ([www.synbiohub.org](http://www.synbiohub.org)).

## 2.2 Synthetic gene circuits in plants

The availability of modular cloning platforms and associated software has eased the design of new gene circuits in plants through the construction and combination of individual functional modules which give plants novel traits and features (Zhu *et al.*, 2021). An early and remarkable example of a complete synthetic circuit engineered in plants are the TNT sentinel plants developed by Antunes *et al.* (2011). In this work, the authors re-designed biological receptors to enable plants to detect specific dangerous substances, responding with rapid and visual output signals. For achieving this, a bacterial periplasmic binding protein (PBPs) involved in chemotaxis was re-designed to detect TNT molecules in the environment and linked to the synthetic histidine kinase-based signal transduction to generate a “de-greening” response consisting of degradation of chlorophyll that generates a bleaching phenotype. This was probably the first attempt to create a full circuit with real-world applications where all the canonical elements of a gene circuit are present, including a TNT receptor (sensor), a chimeric transcriptional activator acting as a processor, and a de-greening actuator system providing a remotely detectable output (Antunes *et al.*, 2006). A more recent synthetic circuit consisted of a flowering control system in rice operated by specific agrochemicals (Okada *et al.*, 2017). In this work, non-flowering rice plants were generated through the overexpression of the floral repressor gene *Ghd7*. This repressor inhibits several floral promotion genes, such as *Hd3a*. The expression of this gene is driven by an agrochemical-inducible promoter rescuing the flowering in the plants a few days after the application of the agrochemical. Unfortunately, at this stage of development of the plant SynBio discipline, the examples of fully implemented functional gene circuits with

agronomic applications are scarce, partially due to the technical difficulty of their implementation beyond model species, but also as a consequence of the limitations imposed on GMO deployment in the field. Despite this, synthetic biology has evolved towards the development of remarkable examples of new synthetic modules, especially sensors, processors, or combinations of both. For testing purposes, the new modules are usually connected to a reporter gene functioning as a minimum actuator (e.g., GUS, GFP), thus providing an easy-to-measure output. These “testing” gene circuits facilitate the evaluation of its components and hopefully will produce the data required for better fine-tuning real-world circuits in the future. The number of new synthetic components is growing rapidly, providing plant biotechnologists with increasingly useful tools with what to undertake new genetic designs (Huang *et al.*, 2021).

### **2.2.1 Sensor modules for plant Synthetic Biology**

In recent years there has been an intense effort in the development of molecular sensors both for endogenous and exogenous signals. Especially interesting as tools for molecular physiological research are hormone sensors. One of the best characterized are the auxin sensors. Initially developed in yeast (Havens *et al.*, 2012), it was later transferred to plants. This sensor, named DII-VENUS, allows the monitoring of the auxin levels in the plant and its distribution through an auxin-triggered degradation of a fluorescent reporter protein (Brunoud *et al.*, 2012). Currently, several optimized versions of the auxin sensors have been developed to respond to different auxin concentrations and offer different visualization outputs, such as the R2D2 and L2min17-Luc sensors based on the degron system (Wend *et al.*, 2013; Liao *et al.*, 2015), the transcriptional sensors DR5:reporter and pIAA motif (Liao *et al.*, 2015; Lieberman-Lazarovich *et al.*, 2019) and the FRET-based AuxSen sensor (Herud-Sikimić *et al.*, 2021). In parallel, a transcriptional cytokinin sensor was also developed, showing GFP patterns that reflect the signalling network of cytokinins in plants. This sensor, apart from its employment in *Arabidopsis*, was correctly adapted to maize (D’Agostino *et al.*, 2000; Zürcher *et al.*, 2013). The jasmonic acid (JA) degradation-based sensors coupled to GFP expression, such as the so-called JAI3-FP (Chini *et al.*, 2007) and Jas9-VENUS (Larrieu *et al.*, 2015), have been developed to visualise the dynamic changes in JA as a stress response in plants. Other

examples of phytohormone sensors are the ABAleons (Waadt *et al.*, 2014), ABACUS (Jones *et al.*, 2014) and SNACS (Zhang *et al.*, 2020) sensors. These FRET-based sensors can generate a change in the fluorescence emission in the presence of ABA (abscisic acid) and monitor levels of this hormone. Also, a great variety of ethylene sensors has been developed to respond to the presence of this phytohormone in plants, such as the degron based EIN3-GFP (Guo and Ecker, 2003) and EIL1-GFP sensors (An *et al.*, 2010) the translational sensors FP-EBF 3'UTR (Merchante *et al.*, 2015) and FP-6x EPU (Li *et al.*, 2015a), and the transcriptional EBS:GUS sensor (Stepanova *et al.*, 2007). Other approaches are focused on the detection of exogenous (Silverstone *et al.*, 2001) or endogenous (Rizza *et al.*, 2017) gibberellin levels in plants, the SA (salicylic acid) visualization (Mou *et al.*, 2003) or the spatiotemporal determination of brassinosteroids (Chaiwanon and Wang, 2015).

Beyond hormone sensors, other types of sensors capable of detecting changes in the environment and generating specific responses are required. Chemical receptors were the first sensors of exogenous stimuli to be adapted to plants, and immediately became essential components of inducible gene expression tools. The tetracycline sensor system, for instance, is a de-repressor of transcription. In absence of tetracycline, the bacterial tetracycline repressor (TetR) binds to the tet operator and blocks the transcription of the coupled gene. This system was adapted to plants and employed for controlling the expression of genes in tobacco, tomato, and potato (Gatz *et al.*, 1992; Weinmann *et al.*, 1994; Bortesi *et al.*, 2012). Contrary to TetR, most chemical sensors act by activating transcription. This is the case of the family of steroid sensors widely employed in plants, which comprise the glucocorticoid sensors (Aoyama and Chua, 1997; Samalova *et al.*, 2005), the estrogen sensors (Bruce *et al.*, 2000; Okuzaki *et al.*, 2011) and ecdysone sensors, which also include the insecticide-induced systems that recognize a synthetic agrochemicals as ligands of the ecdysone receptors from insects (Unger *et al.*, 2002; Koo *et al.*, 2004; You *et al.*, 2006). In these systems, the transcriptional output remains off until the hormone ligand binds to the cytoplasmic steroid receptor allowing its translocation to the nucleus. The steroid receptor is engineered to contain a DNA-binding domain and a regulation domain to induce the transcription of a target gene when it binds to its promoter. Other sensors based on the

specific recognition of small chemical molecules more suitable to be employed in the field were recently developed, such as the copper-induced system or the ethanol-induced system. The induction system based in copper recognizes higher levels of this element than the endogenous levels present in the plant, generating a conformational change of the copper-responsive factor CUP2 fused to an activation domain that results in transcriptional activation of the gene driven by the CBS operon (Saijo and Nagasawa, 2014; Garcia-Perez *et al.*, 2022). The ethanol-induced system is based on the fungal protein AlcR. In presence of ethanol, AlcR binds to the *pAlcA* promoter and generates a transcriptional activation of the coupled target gene (Caddick *et al.*, 1998; Li *et al.*, 2005). In addition, other sensors have been developed through engineering plant receptors to recognize new ligands, such as the engineered ABA receptors that recognize mandipropamid and generate an ABA-related response to control the water uptake in the plant (Rodriguez and Lozano-Juste, 2015).

Other sensors that respond to non-chemical stimuli, such as light, were developed in recent years. Optogenetics sensors are based on cryptochromes or phytochromes, which undergo structural changes when irradiated with a specific wavelength (Christie and Zurbriggen, 2021). In plants, different optogenetic sensors have been developed. The first one to be efficiently employed in *N. benthamiana* and *Arabidopsis* was based on the PHYB-PIF phytochrome interaction that induces transcription by the action of red light (Müller *et al.*, 2014; Ochoa-Fernandez *et al.*, 2016). To perform a tighter control of the gene expression, this approach was combined with the optogenetic system based on the LOV transcription factor (Pudasaini *et al.*, 2015). The generation of a synthetic bipartite promoter controlled by LOV fused to a repressor domain and by PHYB fused to an activator domain leads to a repressed state of the target gene in white light and a specific activation under monochromatic red light (Ochoa-Fernandez *et al.*, 2020). Other optogenetic approaches involve the cryptochrome CRY2-CIB1 sensor (Duan *et al.*, 2017) that transcriptionally activates the target gene by the action of blue light or the CarH photoreceptor system that, in presence of adenosylcobalamin co-factor (AdoB12), activates transcription in the presence of green light (Chatelle *et al.*, 2018). However, this approach is less suitable to work *in vivo* with plants due to the necessity to supplement AdoB12 exogenously. Successful examples of optogenetic circuits in plants

involve, for example, the manipulation of auxin regulatory networks through the red light-inducible system (Müller *et al.*, 2014) and the increased biomass production in *Arabidopsis* achieved through the blue light-induced K<sup>+</sup> channels. This system, named BLINK1, controls stomatal movements and the K<sup>+</sup> uptake in leaves in response to blue light, generating an increase in guard cell volume and turgor, reducing water requirements (Papanatsiou *et al.*, 2019).

The collection of sensors that respond to different stimuli keeps growing. In the context of the development of genetic circuits, sensors represent the first step in the control of gene expression. However, the direct coupling of a sensor to an actuator through a basic processor that offers only the “on” (identify function) or “off” (negation function) transcriptional responses, as it happens in most traditional inducible systems available for plants, is not sufficient to integrate the complex responses required in innovative plant breeding. It is necessary to develop new genetic modules that operate as processors, therefore increasing the range of synthetic transcriptional responses that can be generated in the plant chassis.

### **2.3 Expanding the versatility of gene circuits: processors for gene regulation in plants**

The regulation of transcriptional activity has been imperative in many other fields apart from synthetic biology since the manifestation of phenotypes of interest has been linked to the activation or repression of target genes. In the synthetic biology framework, the elements that process (transfer, amplify, integrate, etc.) the transcriptional signal are categorized as processors, but they have been leveraged long before this discipline emerged.

The classical way to regulate the transcription of a specific set of genes consists in the overexpression of heterologous transcriptional factors (TFs), which usually comprise a DNA binding domain and an effector domain that recruits the transcriptional machinery or other effectors for the control of the gene expression (Zhang, 2003; Hong, 2016). In addition, these TFs could be placed under the control of inducible promoters serving as input sensors, connecting the promoter-specified inputs to a cascade of TF-targeted

activated/repressed genes as outputs (Petolino and Davies, 2013; Li *et al.*, 2013). Nevertheless, a limitation of this approach is that it does not allow free selection of the output response, as the collection of target genes is restricted by the DNA binding specificities of the transcriptional factors employed. An alternative to this approach is the design of artificial promoters in the target genes that included specific binding sites for the selected transcription factor (Kumar *et al.*, 2015; Mohan *et al.*, 2017; Dong *et al.*, 2019; Bai *et al.*, 2020). However, the complexity of the design and the effort required to generate a specific synthetic promoter driving each downstream gene is a factor to consider. Also, this approach may carry a risk of unwanted transcriptional activities.

Target-specific transcriptional interventions not only require optimization of the DNA binding specificity of the effectors, but also the identification and isolation of strong and efficient domains for gene regulation. In general, regulatory domains have been isolated from transcription factors and proteins involved in microbial and viral transcription. One of the first-described transcriptional regulation domains came from the yeast GAL4 TF, whose trans-activation domain (TAD) was identified and isolated (Laughon and Gesteland, 1982). Likewise, the DNA-binding domain of this TF was also isolated and synthetically fused to other potential TADs, which allowed not only the identification of more activation and repression domains, but also demonstrated the orthogonality of the TADs and, therefore, its potential to operate in various species and various genomic contexts (Keegan *et al.*, 1986; Hope and Struhl, 1986). The modular nature of many transcriptional regulation domains led to the identification of powerful viral TADs, such as the VP16 domain of the herpes simplex virus (Campbell *et al.*, 1984; Carey *et al.*, 1990), which proved to be a strong activator also in plants (Schwechheimer *et al.*, 1998; Moore *et al.*, 1998). Additionally, this domain offered the possibility of increasing its transcriptional activation potential through the fusion of several repetitions in tandem, giving rise to the synthetic activation domains VP16, VP64, VP128, etc. These synthetic domains offered a greater activation range of target genes in different biological systems (Ordiz *et al.*, 2002; Li *et al.*, 2017a).

Endogenous regulatory domains have also been identified in plants that can offer either activation or repression. Two examples of plant TADs are the ERF2(m) and the EDLL

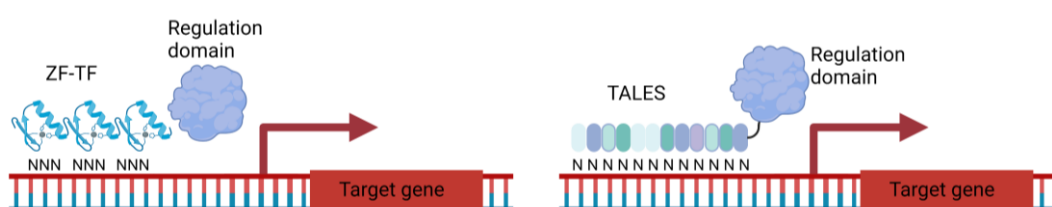


domains, both from the Ethylene Response Factor family (ERF) (Tiwari *et al.* 2012; Li *et al.* 2013). The ERF family includes also proteins with identified transcriptional repression domains, such as the EAR motifs (Ethylene-responsive element-binding factor-associated amphiphilic repression), with which it has been possible to obtain efficient transcriptional repression (Kagale and Rozwadowski, 2011). A remarkable example is ERF3 (Ohta *et al.*, 2001; Tiwari *et al.*, 2004) from which the widely used transcriptional repression domains (TRDs) SRDX and BRD are derived (Hiratsu *et al.*, 2003; Ikeda and Ohme-Takagi, 2009).

Besides activators and repressors operating the simplest identity and negation functions, more sophisticated processors have been developed in plants such as logic gates or toggle switches. Synthetic Boolean logic gates AND, OR, and NOR integrate and combine several inputs to produce one specific output. For example, given the input A and B, the AND logic gate requires both inputs to produce an output. The OR logic gate requires A input, B input, or both to generate an output. Finally, the NOR logic gate only generates an output when none of the inputs are present. The synthetic logic gates have been efficiently adapted in plants for the tight control of gene expression (Shanidze *et al.*, 2020; Belcher *et al.*, 2020; Lee and Woo, 2020). Another important processor type is memory (toggle) switches. These memory systems avoid the need for the continuous addition of external inducers for keeping the output through the punctual detection of the input (Müller *et al.*, 2014). Toggle switches establishing robust bistable equilibrium states in the cells can be found naturally in plants (Schoof *et al.*, 2000; Cruz-Ramírez *et al.*, 2012). Recently, a synthetic reversible memory switch was developed in plants based on the alternative inversion of a central DNA regulatory element composed of a promoter and a terminator in the opposite orientation (Bernabé-Orts *et al.*, 2020). This regulatory element controls the transcriptional activity of two genes of interest, which are located on the sides of the regulatory element, providing a changeable “on” or “off” state of its transcription. Through the action of integrase  $\phi$ C31, the regulatory element is inverted, generating a change in the expression of the genes of interest. The pioneering approach of this switch is that it can return to its initial state through the action of the integrase recombination directionality factor.

### 2.3.1 Programmable transcriptional regulators as versatile elements for the design of genetic processors

Artificial transcription factors consisting of chimeras of DNA-binding domains derived from natural transcriptional factors present as the main drawback of their hardwired DNA specificity, which limits the choice of target genes, as these should obligatorily contain the cognate *cis* DNA operator. The possibility of programming the specificity of transcriptional regulators did arise until the discovery of artificial zinc fingers (ZFs) and transcription activator-like effectors (TALEs). These tools offer the possibility of creating customizable transcriptional factors. The artificial ZFs (Figure 4A) are customizable DNA binding domains that typically recognize 3-6 nucleotide triplets. Initially, artificial ZFs were developed for targeted mutagenesis, producing double-stranded DNA breaks (DBS) upon the fusion of the FokI nuclease (Durai *et al.*, 2005). Later the technology evolves into artificial TFs by including translational fusions to transcriptional regulation domains or epigenetic effectors (Shrestha *et al.*, 2018). Likewise, the engineered TALEs share many similarities in operation and structure with ZFs but offer a higher degree of specificity. TALEs are proteins from bacteria of the genus *Xanthomonas*, which participate in the plant infection mechanisms by promoting the expression of host genes. TALEs (Figure 4B) consist of a specific and customizable DNA binding domain comprising tandem repeat arrays of amino acids which can recognize a specific DNA target sequence (Boch *et al.*, 2009; Moore *et al.*, 2014). As in the case of ZF, the mechanism of TALEs action requires a new design of the protein for each target sequence, which makes them efficient but labour-intensive tools.



**Figure 4: Schematic representation of artificial ZFs (zinc fingers) and TALEs (transcriptional activator-like effectors) acting as programmable transcriptional factors.** The Ns represent the target nucleotide sequences that each technology recognizes.

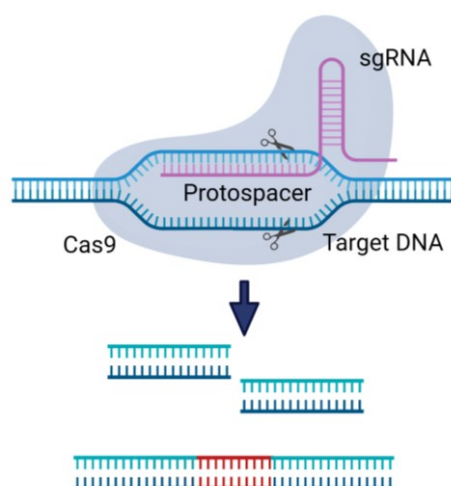
In plants, several works describe ZFs and TALEs being employed as artificial TFs, thus enabling programmable gene regulation. The first ZFs examples targeted the *APETALA3* gene in *Arabidopsis*. The VP64 activation domain and the mSin3 interaction domain (SID) repressor domain were fused to an *APETALA3* ZF yielding the expected transcription changes and generating altered floral patterns (Guan *et al.*, 2002). Later on, engineered TALEs were proven as efficient customizable transcriptional regulators in plants. Interesting examples are the regulation of *EGL3* and *KNAT* endogenous genes in *Arabidopsis* (Morbiter *et al.*, 2010), or the regulation of the *AtPAP1* transgene in tobacco (Liu *et al.*, 2014). More recently, programmable transcriptional regulators based on ZF designs coupled to an epigenetic effector were developed. This approach allowed the demethylation of the *FWA* gene in *Arabidopsis* for controlling the flowering time, using the catalytic domain of human TEN-ELEVEN TRANSLOCATION1 (TET1) (Gallego-Bartolomé *et al.*, 2018), a dioxygenase involved in the demethylation of DNA (Chen *et al.*, 2014).

### **2.3.2 CRISPR tools for gene regulation**

In the last decade, the CRISPR (clustered, regularly interspaced, short, palindromic repeats)/Cas (CRISPR-associated) systems emerged as new versatile programmable effectors. They offer a wide range of applications with high efficiency and specificity, avoiding the main problem that limited previous tools since CRISPR /Cas eliminates the need to make a new protein for each target (Waryah *et al.*, 2018; Arya *et al.*, 2020)

CRISPR/Cas tools are based on ancient immunity systems adopted by some prokaryotic cells and are designed to recognize and eliminate foreign DNA of bacteriophages and plasmids (Mojica *et al.*, 2005). The palindromic repeats delimit variable short spacers, which are transcribed into non-coding RNAs and form a functional complex with CRISPR-associated proteins. The CRISPR/Cas complex is capable of recognizing the complementary invading DNA and generating a double-strand break in the target sequence (Barrangou *et al.*, 2007). Although CRISPR systems were identified from different prokaryotic organisms, they have been classified into two large classes, Class I and Class II, referring to the design principles of the effector module that they contain. Likewise, these two classes are divided into three types each, types I, III and IV in class

1, and types II, V and VI in class 2 (Koonin and Makarova, 2019). It was the CRISPR/Cas9 system of *Streptococcus pyogenes*, belonging to type II, that emerged as the first efficient genomic editing effector that only requires a single Cas associated protein and two non-coding RNA genes to form the active complex, one for trans-activating crRNA (tracrRNA) and other for the precursor crRNA (pre-crRNA). The pre-crRNA, which contains nuclease guide sequences, is processed to mature CRISPR RNA (crRNA) in combination with tracrRNA. However, these elements were reduced to single non-coding RNA genes to form a small guide RNA that contains 20 nucleotides of target sequence (protospacer) and an RNA scaffold which is recognized by the Cas9, emulating the tracrRNA (Jinek *et al.*, 2012; Ran *et al.*, 2013). Furthermore, the binding specificity does not generate off-target problems since it is determined by both, the short DNA motif, named protospacer adjacent motif (PAM) sequence, and the target sequence-DNA base pairing (Haeussler *et al.*, 2016; Figure 5).



**Figure 5: Schematic representation of the CRISPR-Cas9 programmable nuclease.** The small guide of RNA (sgRNA) is represented with its protospacer region. The Cas9 recognizes and cleaves a target DNA region producing NHEJs (Non-homologous end-joining).

The simplicity and versatility of this system based on a small RNA guide (gRNA) and an effector Cas protein overshadow the previous editing tools, proving in its first attempts an efficient genome editing activity in human cells (Cong *et al.*, 2013; Cho *et al.*, 2013; Mali *et al.*, 2013). Moreover, its employment was not just limited to mammalian cells, since its application was extended to all types of organisms (Sanders *et al.*, 2018; Shi *et al.*, 2019; Schuster and Kahmann, 2019; Kim *et al.*, 2019), including the plant kingdom.

In this context, its employment has been proven as an efficient and clean alternative to the traditional methods of mutation and transgenesis, expanding the horizons in crop improvement (Veillet *et al.*, 2019; Gao, 2019; Shao *et al.*, 2020; Zaynab *et al.*, 2020).

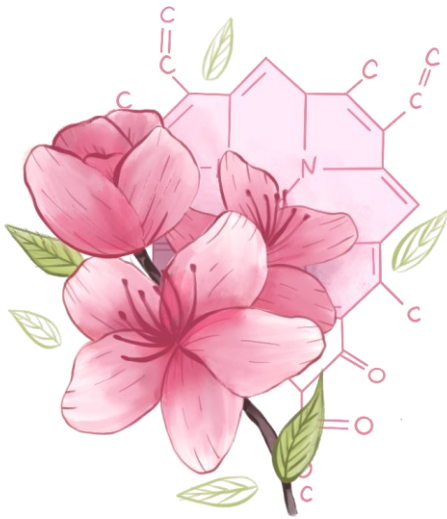
Additionally, the CRISPR system also has the potential to be used as a processor of transcriptional signals. Taking in mind the same strategy employed in ZFs and TALEs, the endonuclease activity of Cas protein was inactivated through the mutation of specific amino acids in RuvC1 and HNH nuclease domains (Qi *et al.*, 2013b). The resulting protein, named “dead Cas” or dCas, can be directed to the gene target promoter and generate a transcriptional response. Depending on the effector domain attached to the dCas protein, specific transcriptional activation or repression of the target gene is achieved (Maeder *et al.*, 2013; Larson *et al.*, 2013). Compared to ZFN and TALEN, CRISPR-based regulators are much easier to program, requiring only the 20 nucleotides protospacer region in the gRNA sequence. At the start of this Thesis, this approach was described with remarkable results in mammalian cells and other organisms, such as bacteria and fungi (Ho *et al.*, 2020; Mózsik *et al.*, 2021), both for targeted activation (also known as CRISPRa strategy) and inhibition/repression (CRISPRi strategy) of gene expression, opening new perspectives for application in plants.

### **2.3.3 Transcriptional regulation with dCas in plants**

The initial strategies to generate programmable transcriptional effectors in plants based on CRISPR-dCas9 employed direct fusions of transcriptional activation or repressor domains to the C-terminus of the Cas9 protein. In plants, the initial CRISPRa approaches encompassed the attachment of TAL, VP64 and EDLL regulation domains to the dCas9 structure (Piatek *et al.*, 2015; Vazquez-Vilar *et al.*, 2016). The results obtained in targeted transcriptional activation of the *AtPAP1*, *AtFIS2*, and *miR319* genes in *Arabidopsis* and the *NbPDS* gene in *N. benthamiana* showed efficient but moderate activation rates (Piatek *et al.*, 2015). Following the same strategy, dCas fusions to the plant-derived BRD and SDRX domains were employed for transcriptional repression of the *NbPDS* gene and a nopaline synthase promoter driving a luciferase reporter in *N. benthamiana* (Piatek *et al.*, 2015; Vazquez-Vilar *et al.*, 2016), showing the same moderate regulatory results. Furthermore, the dCas9 fusion strategy is compatible with attaching epigenetic

regulators to the Cas9 structure, increasing the type of regulation over the target genes. An example in plants is the successful epigenetic regulation of the *AREB1/ABF2* gene in *Arabidopsis* through the direct fusion to dCas9 of the P300 catalytic domain, a domain derived from the *Arabidopsis* HISTONE ACETYLTRANSFERASE1. Although the transcriptional activation of the gene was modest, reaching only 2-fold, a substantial increase in the survival rate was obtained when plants were exposed to drought stress (Roca Paixão *et al.*, 2019).

As the recent works show, CRISPR-Cas9-based TFs developed at the time this thesis started were able to perform specific targeted transcriptional regulation but showed weak activation or repression rates. Nevertheless, these tools had enormous potential as components of genetic processors operating in synthetic gene circuits, since they can be employed for applications requiring a cleavage-free specific DNA binding, such as driving perturbations of native gene networks, controlling endogenous or transgenic metabolic pathways, or generating intricate regulation cascades. In addition, these tools present an unprecedented versatility in the design of the target sequence due to the possibility to generate an almost unlimited battery of gRNAs that can operate through a multiplexing approach to increase the efficiency of the tool. Their modularity and orthogonality bypass the current limitations of traditional TFs, so, an obvious engineering challenge was to maximize its gene regulation capacity for generating specific and strong transcriptional responses and develop efficient genetic circuits with multiple potential outputs.



# Objectives





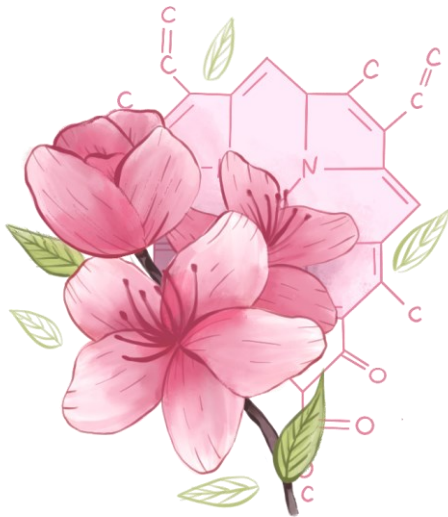
## OBJECTIVES

The main objective of this thesis, framed within the field of plant synthetic biology, is the expansion of the functionalities of genetic circuits in plants, through the development of modular, orthogonal, and versatile CRISPR-dCas9 tools for programmable transcriptional regulation.

To achieve this, the following sub-objectives are proposed:

1. To develop a CRISPR tool to achieve a strong programmable transcriptional activation (CRISPRa) in plants through the comparison of different dCas9 architectures combined with the evaluation of different activation domains.
2. To evaluate the specificity of the new CRISPRa tools through the analysis of the possible activation of off-target genes in the plant transcriptome.
3. To apply the new CRISPRa tools for activating a different set of genes in a plant metabolic pathway, reprogramming the metabolic flux, and obtaining customized metabolic profiles.
4. To expand the range of applications of the new CRISPRa tools by developing new delivery strategies for the CRISPR components.





# Chapter 1

## Development of CRISPR/Cas9-based programmable transcriptional activators for plants

**Selma S**, Bernabé-Orts JM, Vazquez-Vilar M, Diego-Martin B, Ajenjo M, Garcia-Carpintero V, Granell A, Orzaez D. Strong gene activation in plants with genome-wide specificity using a new orthogonal CRISPR/Cas9-based programmable transcriptional activator.

Plant Biotechnology Journal (2019), DOI: 10.1111/pbi.13138

My contribution to this work was essential to this publication. I contributed to most of the analysis performed and a major part of the manuscript writing.



## ABSTRACT

Synthetic Biology (SynBio) aims at rewiring plant metabolic and developmental programs with orthogonal regulatory circuits. This endeavour requires new molecular tools able to interact with endogenous factors in a potent yet at the same time highly specific manner. A promising new class of SynBio tools that could play this function are the synthetic transcriptional activators based on CRISPR/Cas9 architecture, which combine autonomous activation domains (ADs) capable of recruiting the cell's transcription machinery, with the easily customizable DNA-binding activity of nuclease-inactivated Cas9 protein (dCas9), creating so-called Programmable Transcriptional Activators (PTAs). In search for optimized dCas9-PTAs, we performed a combinatorial analysis with seven different ADs arranged in four different protein/RNA architectures. This analysis resulted in the selection of a new dCas9-PTA with improved features as compared with previously reported activators. The new synthetic riboprotein, named dCasEV2.1, combines EDLL and VPR ADs using a multiplexable mutated version (v2.1) of the previously described aptamer-containing guide RNA2.0. We show here that dCasEV2.1 is a strong and wide spectrum activator, displaying variable activation levels depending on the basal activity of the target promoter. Maximum activation rates reaching up to 10000-fold were observed when targeting the *NbDFR* gene. Most remarkably, RNAseq analysis of dCasEV2.1-transformed *N. benthamiana* leaves revealed that the topmost activation capacity of dCasEV2.1 on target genes is accompanied by strict genome-wide specificity, making dCasEV2.1 an attractive tool for rewiring plant metabolism and regulatory networks.

**Keywords:** CRISPR/Cas9, dCas9, scRNA, transcriptional activation, synthetic biology, GoldenBraid

## INTRODUCTION

The ability to superimpose synthetic regulatory circuits on plant endogenous gene expression networks has been aimed for a long time as it will open a range of applications in plant biotechnology. Classical attempts have involved the ectopic expression of heterologous transcriptional factors (TFs) under the control of purpose-specific promoters, therefore connecting promoter-specified inputs to a cascade of TF-targeted activated/repressed genes as output (Petolino, 2015). An important limitation of this approach is that it does not allow free selection of the output response, as the collection of target genes is restricted by the DNA binding specificities of the TFs employed, which are typically “hardwired” at the protein level. A way to circumvent this limitation is to engineer artificial promoters with TF-binding cis-regulatory elements controlling all target genes (Venter, 2007; Roccaro *et al.*, 2013; Kumar *et al.*, 2015). However, this requires important engineering efforts, limiting its applicability in practical terms. Furthermore, “natural” TFs have often a broad spectrum of DNA binding activities, limiting the orthogonality and specificity of the approach.

Contrary to hardwired TFs, CRISPR/Cas9 protein architecture enables the design of transcription factors with DNA-binding specificities that can be programmed in a 20 nucleotide-long guide RNA with minimum engineering efforts. Using catalytically inactive Cas9 (dCas9) nuclease fused to transcriptional activator or repressor domains, the gene expression of an individual gene can be modified (Maeder *et al.*, 2013). The CRISPR-dCas9 strategy presents advantages as compared with previously described modifiable regulators including zinc finger nucleases, ZFNs (Steger *et al.*, 2002; Lindhout *et al.*, 2006), or TAL effectors (Liu *et al.*, 2014; Gao *et al.*, 2014) which are very specific but require a new recoding of regulatory protein for each target sequence.

Initial attempts to produce CRISPR-dCas9 Programmable Transcriptional Activators (Cas9-PTAs) in plants made use of transcriptional activator domains (TAD) fused to dCas9 protein, achieving only low/moderate activation rates (Vazquez-Vilar *et al.*, 2016; Park *et al.*, 2017). The next generation of PTAs are designed to achieve increased activation rates by combining several TADs displayed in a single dCas9 protein. To do so,

different strategies have been proposed. The SunTag strategy (Tanenbaum *et al.*, 2014b) uses multi-epitope tags to attach multiple TADs, whereas SAM and CRISPR scaffold RNAs (scRNA) strategies (Konermann *et al.*, 2015a; Zalatan *et al.*, 2015; Zhang *et al.*, 2015c) use RNA aptamers added to the gRNA scaffold as secondary anchoring sites for TADs. All those strategies have been showed to improve activation rates in mammalian cells and recently shown similar results in plants (Lowder *et al.*, 2015, 2018a). Despite these achievements, the search for optimal PTAs in plants is still open to improved designs. Ideally, plant PTAs should combine strong activation rates with a wide spectrum of responsive targets and most importantly, with high target specificity, that is, ensuring that the cell transcriptome is only affected in the intended gene(s).

In a search for improved plant dCas9-PTAs, here we show the results of a systematic comparison of 43 SunTag, SAM and scRNA-based TAD combinations tested for their ability to activate different promoters fused to a Luciferase reporter. As a result of this analysis, we selected a new dCas9-PTA comprising two TADs (EDLL and VPR) in combination with an aptamer variant of the gRNA2.0 scaffold employed in the scRNA strategy. The new dCas9-PTA (named dCasEV2.1) consistently produced the highest activation rates in all assays, both using transiently-transformed and genome-integrated promoters as targets. dCasEV2.1 was able to activate the *Nicotiana benthamiana* Dihydroflavonol-4-reductase (DFR), an inducible gene with very low basal activity levels, up to 10000 folds. When directed to strong constitutive promoters, dCasEV2.1 yielded lower induction rates but raised transcriptional activity up to levels that triplicate those of the strong CaMV35S promoter. Moreover, RNAseq analysis showed a remarkable genome-wide specificity of dCasEV2.1-PTA, with virtually no changes in the transcriptome other than those anticipated by off-target analysis.

## **MATERIALS AND METHODS**

### **GB phyto bricks construction and assembly**

Level 0 plasmids used in this work were created following the domestication strategy described in Sarrion-Perdigones *et al.* (2013), Plasmids pHRdSV40-dCas9-24xGCN4-v4-

P2A-BFP, Addgene ID: #60903 and pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS, Addgene ID #60904 (Tanenbaum *et al.*, 2014a), kindly provided by Ron Vale laboratory, served as a template for the construction of GB2464 and GB1463 by PCR amplification using the Phusion High Fidelity DNA polymerase. All level 0 parts are listed in Supplementary Table 1 and their sequences can be searched at <https://gbcloning.upv.es> with the corresponding GB IDs. Multipartite Bsal restriction-ligation reactions from level 0 parts and binary Bsal or BsmBI restriction-ligation reactions were performed as described (Sarrion-Perdigones *et al.*, 2013) to obtain all the level  $\geq 1$  assemblies. A list of all the TUs and modules used in this work is provided in Supplementary Table 2. All level  $\geq 1$  constructs were validated by restriction enzyme analysis. All gRNAs used in this work were designed using the Benchling CRISPR tool ([www.benchling.com](http://www.benchling.com)) following the scheme described in Supplementary Figure 1A for gRNA position determination. For single gRNA assembly in GB level 1 plasmid (Supplementary Figure 1B), primers including the protospacer sequence were designed at [www.gbcloning.upv.es/do/crispr](http://www.gbcloning.upv.es/do/crispr). All primers used for gRNAs assembly are listed in Supplementary Table 3. Primers were resuspended in water to final concentrations of 10  $\mu\text{M}$ . Equal volumes of forward and reverse primers for each gRNA were mixed. The mixture was incubated at room temperature for 5 min for the hybridization of the primer pair. gRNA assembly in level 1 vector was carried out with a Bsal restriction-ligation reaction. The reactions were set up in 10  $\mu\text{l}$  with 1  $\mu\text{l}$  of primers mix, 75 ng of U626 promoter (GB1001), 75 ng of the corresponding scaffold (GB0645, GB1436, GB2461, GB1437, GB1450 or GB1451) and 75 ng of pDGB3 $\alpha$  destination vector. All gRNAs in level 1 plasmids used in this work are listed in Supplementary Table 4. For the assembly of gRNAs to be used in the multiplexing strategy, GB level -1 plasmids containing the tRNA and the gRNA2.1 scaffold were designed following the plasmid structure described in Vazquez-Vilar *et al.* (2016). For the construction of these plasmids, DNA fragments including the tRNA and the gRNA2.1 scaffold were synthesized as IDT gBlocks<sup>®</sup> Gene Fragments and subsequently cloned in pVD1 (GB0101) with a Bsal restriction-ligation reaction. Level -1 plasmids are listed in Supplementary Table 5. Individual gRNAs were assembled in pUPD2 with a BsmBI restriction-ligation reaction that was performed with 75 ng of pUPD2 and 75 ng of the corresponding level -1 pVD1: pVD1\_M1-3pTRNA scf 2.1, pVD1\_M2-3pTRNA scf



2.1 or pVD1\_M3-3pTRNA scf 2.1 plasmid, depending on the desired position of each target in the polycistronic gRNA, and a mix of complementary primers with the protospacer sequence. Supplementary Figure 1C shows the schematic representation of the multiplexing gRNAs cloning strategy. All level 0 gRNAs generated with this cloning strategy are listed in Supplementary Table 6. All constructs were validated by restriction enzyme analysis and confirmed by sequencing.

### ***Nicotiana benthamiana* agroinfiltration**

The transient expression assays were carried out through agroinfiltration of *N. benthamiana* leaves. *N. benthamiana* plants were grown for 5 weeks before agroinfiltration in a growth chamber where the growth conditions were 24°C/20°C light/darkness with a 16h/8h photoperiod. The plasmids were transferred to *Agrobacterium tumefaciens* strain GV3101 by electroporation. Agroinfiltration was carried out with overnight-grown bacterial cultures. The cultures were pelleted and resuspended in an agroinfiltration solution (10 mM MES, pH 5.6, 10 mM MgCl<sub>2</sub>, and 200 μM acetosyringone). After incubation for 2 h at room temperature with agitation, the optical density of the bacterial cultures was adjusted to 0.1 at 600 nm. Cultures were mixed in equal volumes for co-infiltration. The silencing suppressor *P19* (GB1203) was included in all the assays; in the same T-DNA for the transcriptional regulation experiments with the reporter constructs and co-delivered as an independent T-DNA for the transcriptional regulation of endogenous genes assays. Agroinfiltrations were carried out through the abaxial surface of the three youngest fully expanded leaves of each *N. benthamiana* plant with a 1 ml needle-free syringe.

### **Luciferase/renilla activity determination**

The assay conditions follow the experimental standards found in [https://gbcloning.upv.es/add/experiment/SE\\_002](https://gbcloning.upv.es/add/experiment/SE_002) with minor modifications. Samples were collected at 5 days post infiltration (dpi) instead of 4 dpi. For the determination of the luciferase/renilla (Fluc/Rluc) activity, one disc per leaf (d = 0.8 cm, approximately 18–19 mg) was excised with a hole puncher, homogenized and extracted with 375 μl of 'Passive Lysis Buffer,' followed by 10 min of centrifugation (14,000×g) at 4 °C. Then, the

supernatant was collected as working plant extract. Fluc and Rluc activities were determined following the Dual-Glo<sup>®</sup> Luciferase Assay System (Promega) manufacturer's protocol with minor modifications: 7.5  $\mu$ l of working plant extract, 30  $\mu$ l of LARII and 30  $\mu$ l of Stop&Glow Reagent were used. Measurements were made using a GloMax 96 Microplate Luminometer (Promega) with a 2-s delay and a 10-s measurement. Fluc/Rluc ratios (RPU) were determined as the mean value of three biological replicates coming from three independent agroinfiltrated leaves of the same plant and were normalized to the Fluc/Rluc ratio obtained for a reference sample that measures the basal activity of the evaluated promoter fused to the reporter.

## **Generation and selection of *N. benthamiana* reporter lines**

The *N. benthamiana* reporter *pNos*, *SIDFR* and *SIMTB* lines were generated following the transformation protocol described previously (Clemente, 2006). Constructs GB2248 (*pNos*), GB2250 (*SIDFR*), GB2249 (*pMTB*) were transferred to LBA4404 *Agrobacterium tumefaciens* strain use for plant transformation. Murashige and Skoog (MS) plates supplied with Kanamycin at 100 mg/ml were used to select the transgenic T0 lines. T1 plants were selected for single-copy T-DNA insertions based on the selectable marker segregation of their offspring. Single-copy T2 lines were sorted for further analysis based on their Fluc/Rluc activity rates. For *pNos*, three homozygous T2 lines representing low (Nos-RL6), medium (Nos-RL3), and high (Nos-RL5) Fluc/Rluc expression levels were selected. For *SIMTB*, analysis was conducted using the T2 homozygous line showing higher expression levels (MTB-RL3). Finally, the T2 heterozygous DFR-RL1 reporter line was selected based on its high Fluc/Rluc induction rates when agroinfiltrated with an ANT1 TF (Mathews *et al.*, 2003). All selected lines were grown under 24°C/20°C light/darkness with a 16h/8h photoperiod condition in a growth chamber.

## **RNA isolation and RT-qPCR gene expression analysis**

Total RNA was isolated from 100 mg of fresh leaf tissue harvested 4 and 7 dpi using a Macherey-Nagel RNA isolation kit according to the manufacturer's instructions. Before cDNA synthesis, total RNA was treated with rDNAse-I Invitrogen Kit according to the manufacturer's recommendations. An aliquot of 1  $\mu$ g of total RNA was used for cDNA

synthesis using PrimeScript™ RT-PCR Kit (Takara) in 20 µl final volume according to the manufacturer's protocols. Expression levels of each gene were measured in triplicated reactions, performed with the same cDNA pool for each condition, in the presence of fluorescent dye (SYBR® Premix Ex Taq) using Applied biosystem 7500 Fast Real-Time PCR system with specific primer pairs (Supplementary Table 7). *F-BOX* protein gene was used as an internal reference (Liu *et al.*, 2012). Basal expression levels were calculated either with samples agroinfiltrated with *P19* and samples with the dCas9-activation TUs without gRNAs. Calculations of each sample were carried out according to the comparative  $\Delta\Delta CT$  method (Livak and Schmittgen, 2001).

## RNA sequencing and analysis

RNA samples were collected and isolated following the protocol for RNA isolation described above. Three biological replicates were selected for each condition: control condition, *NbDFR* activation and *NbAN2* activation. The control condition sample was agroinfiltrated with the T-DNA that contains the *dCasEV2.1* construct without any gRNA. RNA-sequencing was undertaken using TruSeq Stranded mRNA LT Sample Prep Kit for library construction. Libraries were sequenced using Illumina HiSeq 4000 platform. 40 million 150 bp paired-end reads per sample were generated and mapped using *N. benthamiana* genome Niben v1.0.1 as a reference, available at [www.solgenomics.net](http://www.solgenomics.net), through the Hisat2 program (Kim *et al.*, 2015). The quality of the reads obtained was evaluated with the FastQC program available online at: [www.bioinformatics.babraham.ac.uk/projects/fastqc](http://www.bioinformatics.babraham.ac.uk/projects/fastqc). The Illumina adaptors were eliminated from the reads using the program Trimmomatic (Bolger *et al.*, 2014). For each sample, a count of the expression was performed using the program StringTie (Pertea *et al.*, 2015), following the gene models for the genome reference Niben v1.0.1, also available on [www.solgenomics.net](http://www.solgenomics.net). Statistical analysis with the raw data was performed to evaluate the differential expression between the samples *NbDFR* and *NbAN2* against control samples, using a statistical package edgeR (Robinson *et al.*, 2010). For each comparison of the expression data, the results were organized depending on the profile of the expression based on the up-regulation or down-regulation of the genes, and GO term analysis was also performed using Blast2GO (Götz *et al.*, 2008) with FDR<0.05. The

off-target analysis was performed through Benchling ([www.benchling.com](http://www.benchling.com)) and Cas-OFFinder (Bae *et al.*, 2014) allowing up to 4 nucleotides of mismatch. Possible off-targets located within 1000bp upstream and 500bp downstream the TSS of the genes were analysed to see if they were differentially expressed (Quinlan and Hall, 2010). The reads have been deposited in the Sequence Read Archive (SRA), under the Bioproject PRJNA507084 ([www.ncbi.nlm.nih.gov/bioproject/507084](http://www.ncbi.nlm.nih.gov/bioproject/507084)).

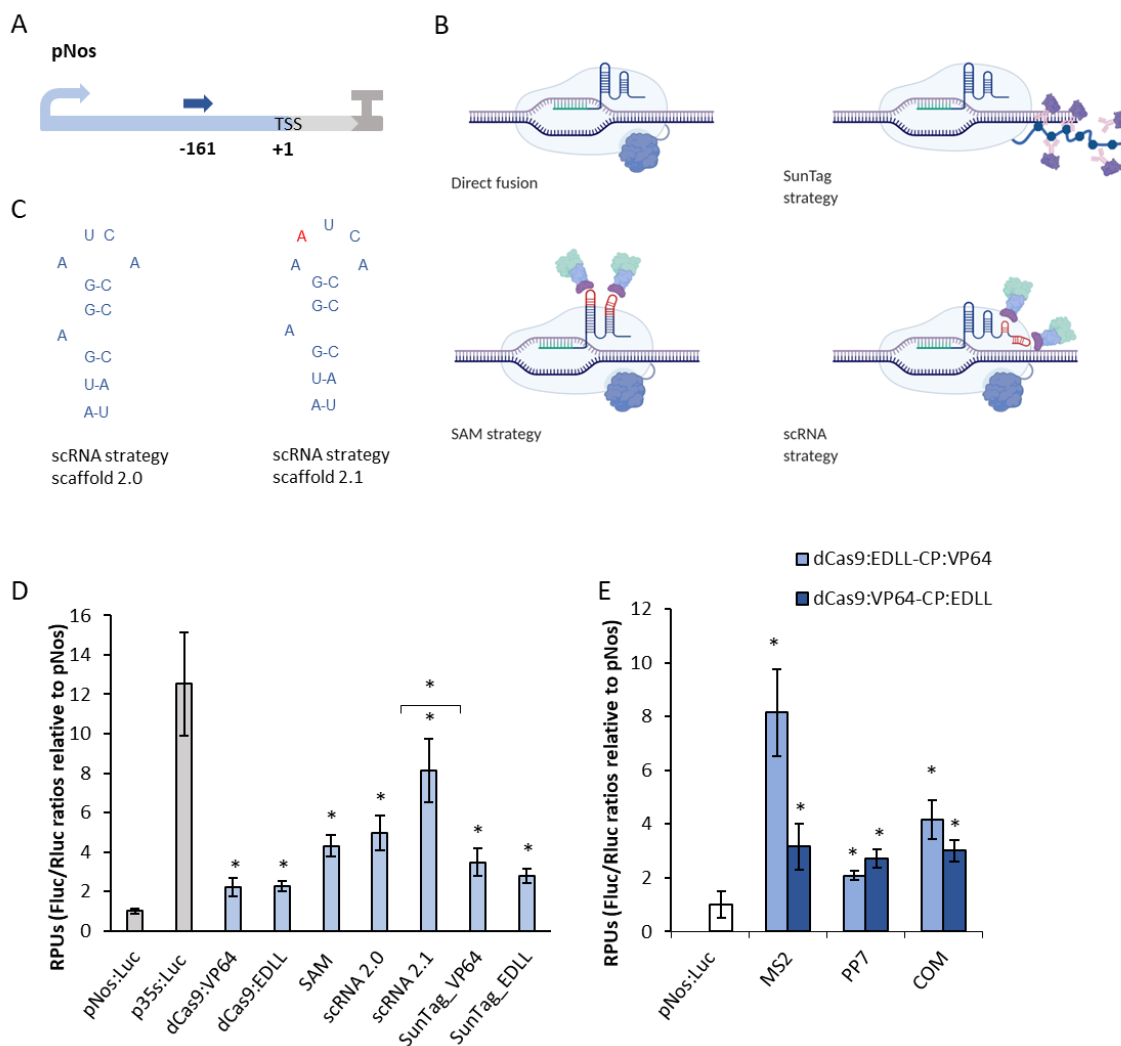
## RESULTS

### CRISPR/Cas9-based Programmable Transcriptional Activator optimization

The first step for developing a robust transcriptional activation tool consists of a round of comparisons among the different PTA dCas9-strategies performed transiently in *N. benthamiana* leaves. PTA transactivation levels were assessed using *Nopaline synthase* promoter (*pNos*) coupled to firefly *luciferase* (Fluc) reporter. A constitutive renilla *luciferase* (Rluc) was used as an internal reference, driven by the CaMV35S promoter. For initial comparisons, PTAs were targeted to the *pNos* promoter with a single gRNA annealing at position -161 relative to the Transcriptional Start Site (TSS) (Figure 1A). This position was validated in previous experiments (Vazquez-Vilar *et al.*, 2016). SAM, scRNA and SunTag (Figure 1B) designs were analysed using the activation domains VP64 and EDLL. For SunTag, VP64 and EDLL were fused to the single-chain fragment variable (ScFv) antibody and tested separately. For scRNA and SAM, EDLL was attached to dCas9, and VP64 was fused to the MS2 viral coat protein, which binds an RNA aptamer. RNA scaffolds in scRNA and SAM designs contained a second optimized aptamer next to the wild-type one, as this double-aptamer design was earlier found to improve binding activity (Nowak *et al.* 2016). During the cloning of the gRNA2.0 scaffold employed in the scRNA design, a spontaneous mutation occurred consisting of the insertion of an adenine in the loop of the first aptamer (Figure 1C). Since this aptamer variant had not been studied earlier, we decided to include it in the comparison analysis (labelled as gRNA2.1). The experiment was completed with direct dCas9:VP64 and dCas9:EDLL

fusions used to monitor the improvement achieved from earlier designs. Relative transcriptional activities (RTAs) in transient assays were expressed as reference promoter units (RPUs). RPUs are calculated as the Fluc/Rluc ratios measured in each sample, normalized with the Fluc/Rluc ratios produced by an unactivated *pNos* promoter (GB1398) assayed in parallel. This procedure was earlier proposed as a standard measurement for the documentation of standard DNA parts (PhytoBricks) (Patron *et al.*, 2015; Vazquez-vilar *et al.*, 2017). In these conditions, the *Cauliflower mosaic virus* 35S (CaMV35S) promoter-based standard DNA part GB0164 consistently confers activity levels of  $12 \pm 2$  RPUs. Thus, this CaMV35S reporter was included in all comparative experiments to serve as an upper limit reference.

As shown in Figure 1D, scRNA, SAM and SunTag strategies produced higher activation rates as compared with previous simplified designs of EDLL or VP64 dCas9 fusions (Vazquez-Vilar *et al.*, 2016). Furthermore, designs involving combinations of different TADs (scRNA and SAM) showed stronger transactivation than the SunTag design, which involves multiple copies of the same TAD. Surprisingly, the highest transcriptional activation was achieved with the scRNA-gRNA2.1 scaffold (scRNA2.1), reaching RTA levels close to that of CaMV35S (RTA =  $8 \pm 2$  RPUs) and outperforming the previously optimized scRNA-gRNA2.0 strategy. In a separate experiment, the gRNA2.1 was compared with similar aptamer-binding coat proteins (CP) PP7 and COM (Zalatan *et al.*, 2015) using the equivalent TAD combinations dCas9:EDLL-CP:VP64 and dCas9:VP64-CP:EDLL (Figure 1E). Again, the gRNA2.1 showed the best activation results. Note that by swapping the position of the activation domains, different levels of activation were achieved.



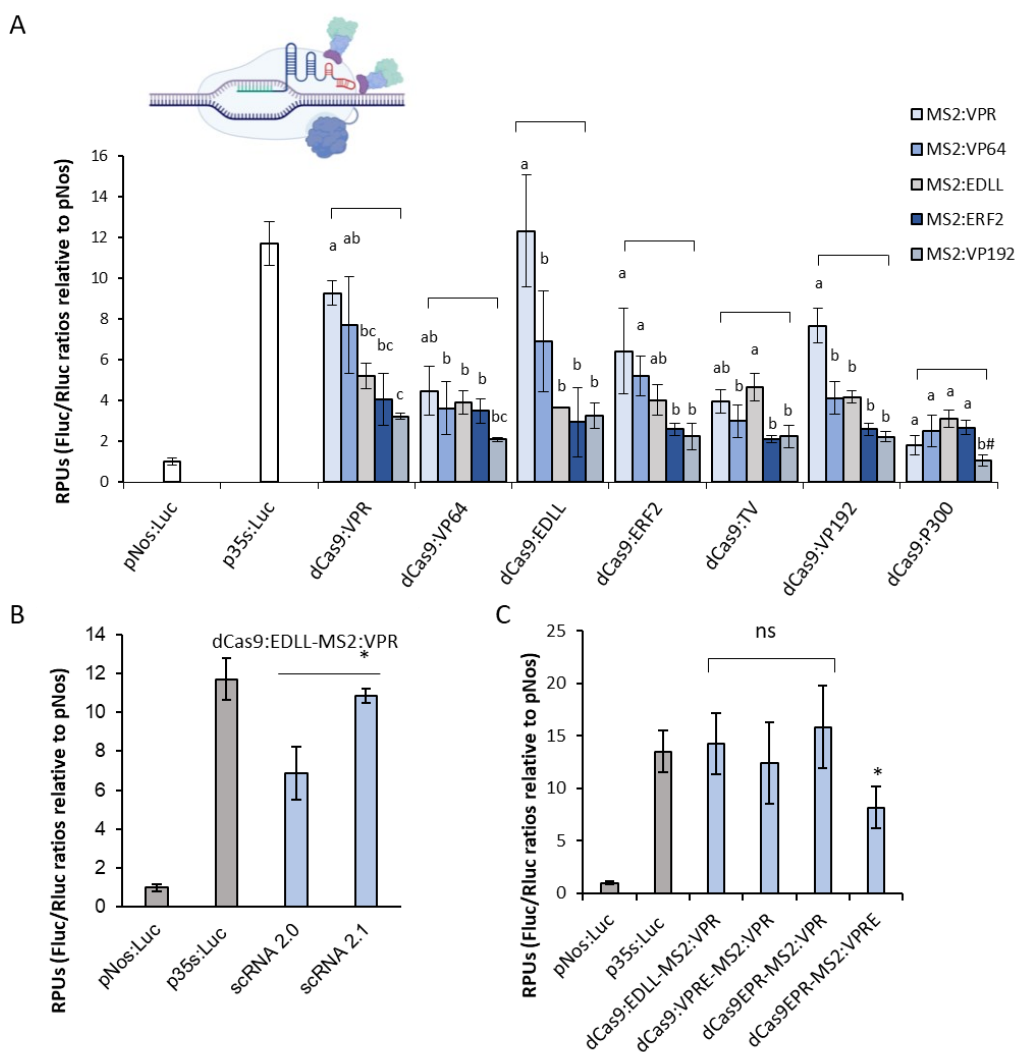
**Figure 1: Transient activation of pNos promoter by dCas9-PTAs. Screening for the best strategy of dCas9-PTAs.** A) Representation of pNos promoter with a gRNA target at position -161. B) Schematic representation of SunTag strategy, scRNA strategy (with gRNA2.0 or gRNA 2.1), and SAM strategy used in activation experiments. C) Representation of the structure of MS2 aptamer loops from the scaffold 2.0 and the mutated version of scaffold 2.1. The red letter represents the inserted adenine in the sequence. D) Relative transcriptional activities (RTAs) obtained with different dCas9-PTAs strategies. All SAM and scRNA strategies combine direct dCas9-EDLL fusions with gRNA aptamers attached to MS2-VP64. RTAs of pNos and p35S reporters measured in the same example are also included as a reference. (E) RTAs obtained upon activation of pNos at -161 position with different aptamer-coat protein combinations using scRNA architecture. MS2 strategy combines scRNA with gRNA2.1; PP7 and COM strategies use a single aptamer to bind PP7 and COM coat proteins respectively; coloured bars represent the two different TAD combinations assayed in each strategy. RTAs are measured as relative promoter units (RPUs) calculated as the Fluc/Rluc ratios for each construct normalized with the Fluc/Rluc ratio of a pNos:Luc

reporter transformed in parallel. The pNos:Luc measure is a reference sample with a dCasEV2.1 and a non-specific gRNA. Asterisks indicate Student's-test significant values ( $p < 0.05$ ). Bars represent average RTAs  $\pm$  SD,  $n=3$ .

In a further optimization step and attending to the differences obtained in TAD swapping experiments described above, we tested the scRNA2.1 design with new TAD combinations as shown in Figure 2A. Seven previously described TADs (VP64, EDLL, P300, VPR, VP192, ERF2, and TV) were fused to dCas9 and combined with five of them (VP64, EDLL, VPR, VP192 and ERF2) anchored at the MS2-aptamer position. A detailed description of each of the TADs employed in this experiment is shown in Table 1. VPR domain showed the highest activation rates in most combinations assayed, followed by VP64 and EDLL. Remarkably, the dCas9:EDLL-MS2:VPR gave maximum *pNos* activation up to RTA levels similar to that of CaMV35S (GB0164) used as our upper limit reference. A separate experiment shown in Figure 2B confirmed that, as expected, the selected dCas9:EDLL-MS2:VPR tandem performed better when bound to the newly described gRNA2.1 aptamer.

**Table 1. Description of TADs employed in combinatorial assays**

TAD	Description	Ref.
<b>VP64</b>	A tetrameric repeat of herpes simplex VP16.	(Beerli <i>et al.</i> , 1998)
<b>EDLL</b>	Short motif present in AP2 sub-family. Activation domain.	(Tiwari <i>et al.</i> , 2012)
<b>P300</b>	p300 core. Acetylase of histone H3 lysine 27.	(Hilton <i>et al.</i> , 2015)
<b>VPR</b>	(VP64-p65-Rta) Tripartite transactivation domain. p65 is a subunit from the NF-kappa B transcription factor that contains the trans-activation domain. Rta is an Epstein-Barr virus R transactivator.	(Chavez <i>et al.</i> , 2015)
<b>VP192</b>	A trimeric repeat of VP64.	(Balboa <i>et al.</i> , 2015)
<b>ERF2(m)</b>	The modified trans-activation domain of ERF2 protein.	(Li <i>et al.</i> , 2013)
<b>TV</b>	6xTAL –VP128.	(Li <i>et al.</i> , 2017a)



**Figure 2: Screening of the TADs combinations employing the scRNA 2.1 strategy through the transient activation of the pNos promoter.** A) RTAs obtained upon activation with different combinations of TADs using scRNA 2.1 strategy, targeting the reporter pNos:Luc at position -161. ANOVA test was performed for each dCas9-PTA group. Bars sharing the same letter are not significantly different according to Tukey's HSD test ( $p < 0.05$ ). The # symbol indicates no statistical differences with the control sample. B) Comparison of RTAs obtained upon the activation with the scRNA dCas9:EDLL-MS2:VPR strategy employing the gRNA 2.0 and 2.1, targeting the reporter pNos:Luc at position -161. Asterisks indicate Student's-test significant values ( $p < 0.05$ ). C) RTAs obtained upon activation with different combinations of VPR and EDLL fusions using scRNA 2.1 strategy, targeting the reporter pNos:Luc at position -161. RTAs are measured as relative promoter units (RPUs) calculated as the Fluc/Rluc ratios for each construct normalized with the Fluc/Rluc ratio of a pNos:Luc reporter transformed in parallel. The pNos:Luc measure is a reference sample with a



---

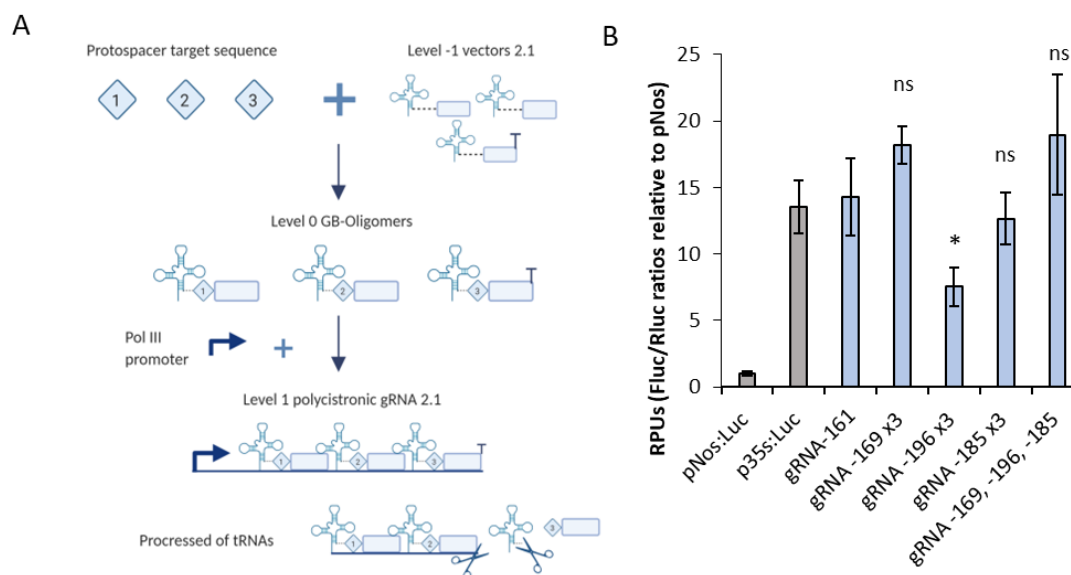
dCasEV2.1 and a non-specific gRNA. Each bracket indicates the group of samples analysed with the corresponding statistical method. Bars represent average RTAs  $\pm$  SD, n=3.

However, given the evidence that the tandem fusion of domains carried out by the VPR (VP64-P65-RTA) domain offered an increase in the activation rates obtained with the scRNA2.1 system, new combinations of domains were generated based on this same strategy. The EDLL and VPR activation domains were selected for this purpose due to their combination providing the best transcriptional activation (Figure 2A). The EPR (EDLL-P65-RTA) and VPRE (VP64-P65-RTA-EDLL) tandem fusions were generated. New combinations of dCas9:AD and MS2:AD were generated by employing these tandem activation domains and the VPR domain and tested following the same strategy described above for targeting the *pNos* promoter. Furthermore, the dCas9:EDLL-MS2:VPR combination was included as a fair comparison. Unfortunately, as is shown in Figure 2C, none of the new combinations showed a significant improvement, compared to the original version based on dCas9:EDLL-MS2:VPR. Surprisingly, the combination based on dCas9:EPR-MS2:VPRE showed a decrease in transcriptional activation, indicating that the fusion of MS2:VPR is a critical component for reaching the best results with the scRNA strategy.

## Multiplexing gRNA cloning for scRNA2.1 strategy

Although obtaining a robust transcriptional activation tool is the main objective of this work, it is also necessary to obtain multiplexing systems for gRNAs based on GoldenBraid adapted to the new requirements of the components of the scRNA strategy. The current gRNA assembly system presents some limitations, such as the requirement of the initial G or the impossibility of generating multiple gRNAs from a single transcript (Xing *et al.*, 2014). In order to solve this problem and maximize the potential of the new transcriptional activation tools, a GB polycistronic design based on the scRNA2.1 strategy was developed. Taking as reference the available multiplexing tool described in Vazquez Vilar *et al.* (2016), a gRNA multiplexing assembly scRNA2.1 for three positions was designed, generating three new level -1 vectors that include, in each of them, a tRNA sequence and the scRNA2.1 scaffold flanked by BsmBI sites (Figure 3A). The protospacer sequence is included in level 0 GB-oligomer through a BsmBI assembly of

the corresponding level -1 plasmid. Finally, these resulting level 0 parts are combined together with the PolIII promoter to create a level 1 polycistronic tRNA-gRNA2.1. The transcript is later processed by the endogenous tRNA ribonucleases P and Z to produce the individual gRNAs (Xing *et al.*, 2014; Figure 3A). In order to test the scRNA2.1 multiplexing design, three new positions of the pNos promoter were targeted. Taking the previously assayed position -161 as a reference, new gRNAs targeting -169 and -196 in the non-coding strand and -185 in the coding strand of the pNos promoter were selected and assembled, individually and in combination. The multiplexing scRNA2.1 test was performed on the *pNos* promoter following the same Fluc/Rluc reporter methodology as described above for the dCas9:EDLL-MS2:VPR combination. Also, the well-studied gRNA-161 was included as a reference. The results obtained using the scRNA2.1 multiplexing system show efficient activation rates (Figure 3B), using a single target or a combination of them. Surprisingly, these results suggest that the effective distance to the gRNA target site from the TSS for mediating dCas9-AD-mediated transcriptional activation may be limited. Moving the target sequence by a few nucleotides further upstream from the TSS can result in a relevant decrease in the activation, as shown in the data obtained with the target -196, which results in 2-fold lower pNos activation, compared with the -169 targeted position. Alternatively, the decrease in the pNos activation seen with gRNA-196 can be the consequence of the less optimal gRNA target sequence. Furthermore, the combination of the three positions does not significantly improve the transcriptional activation obtained, leading us to conclude that it is possible that the maximum transcriptional activity allowed by the promoter has been reached.



**Figure 3: Design and testing of multiplexed gRNA cloning for scRNA2.1 strategy.** A) Schematic representation of the multiplexing strategy adapted to GoldenBraid assembly system incorporating the gRNA 2.1 sequences. B) RTAs obtained upon activation with the multiplexing vectors and the scRNA 2.1 strategy targeting the reporter pNos:Luc at positions -169, -196, and -185. The gRNA that targets the *pNos* promoter in -161 was included for fair comparisons. RTAs are measured as relative promoter units (RPUs) calculated as the Fluc/Rluc ratios for each construct normalized with the Fluc/Rluc ratio of a pNos:Luc reporter transformed in parallel. The pNos:Luc measure is a reference sample with a dCasEV2.1 and a non-specific gRNA. The Student's test was performed for comparing the values of Rluc/Fluc obtained with the new multiplexing vectors (gRNA -169 x3, gRNA -196 x3, gRNA -185 x3, gRNA -169,-196,-185) with the activation reference (gRNA -161). Asterisks indicate Student's-test significant values ( $p < 0.05$ ). Bars represent average RTAs  $\pm$  SD,  $n=3$ .

## dCasEV2.1, an efficient tool for a large range of targets

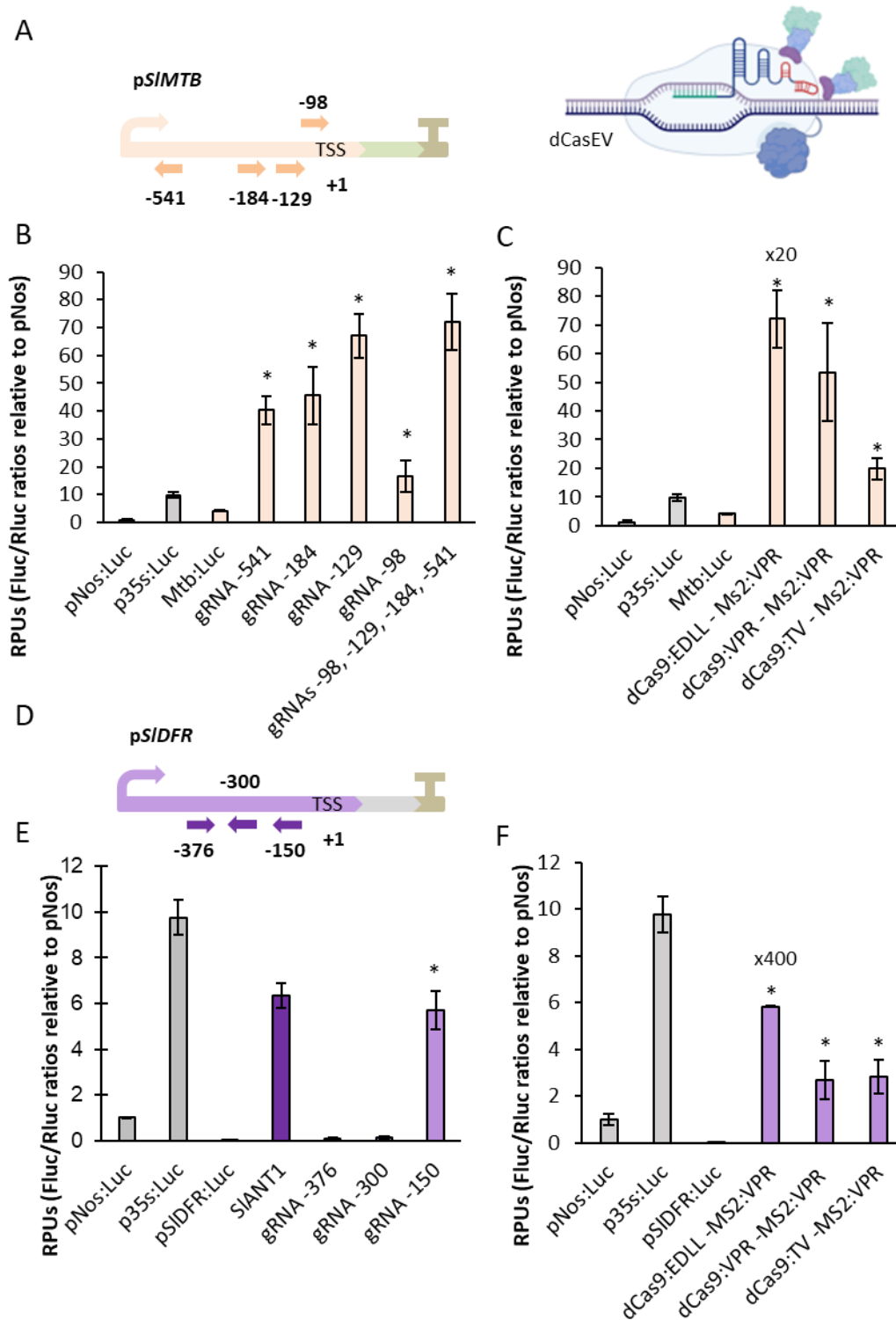
The results obtained with the pNos:Luc reporter prompted us to test the activity of dCas9:EDLL-MS2:VPR/gRNA2.1 (now abbreviated as dCasEV2.1) on promoters with contrasting basal activities. The promoter derived from the tomato *SIMTB* (*Metallothionein-like protein type2B*) gene (catalogued as GB1399), has a strong constitutive basal activity (RTA=  $4 \pm 1$  RPUs) and could serve to test dCasEV2.1 ability to rise transcription levels in absolute terms. On the contrary, the promoter of the *SIDFR*

gene (*Dihydroflavonol-4-reductase*) from *Solanum lycopersicum* (*pSIDFR*, catalogued as GB1160), has very low basal expression levels (RTA < 0.04 RPU). This gene belongs to the flavonoid pathway that is repressed in *N. benthamiana* (Outchkourov *et al.*, 2014), but it is strongly induced in planta by the presence of “natural” MYB TFs (e.g. *SIANT1*). *pSIDFR* was used as a model promoter to test the activation range of dCasEV2.1 induction, since *N. benthamiana* does not express any MYB that could mask the CRISPR activation over this promoter.

Activation of the *pSIMTB* was first analysed using dCasEV2.1 complex and gRNAs at positions -98, -129, -184 and -541, represented in Figure 4A. The gRNAs were tested individually or combined in a single T-DNA with the GoldenBraid multiplexing cloning strategy. As observed in Figure 4B, all gRNA tested in a range of 500 bp upstream of the TSS conferred strong activation to the promoter, with position -129 reaching maximum levels. The combination of gRNAs, in this case, did not confer activation levels higher than those obtained by gRNAs acting individually. Most notably, absolute RTA levels obtained with gRNA -129 reached record RTA levels ( $60 \pm 10$  RPU), corresponding to a 20-fold activation relative to the *pSIMTB* basal levels and 4 times above the CaMV35 levels used as an upper limit reference. The superior performance of the dCasEV2.1 complex was once more confirmed in a separate activation experiment with gRNAs in positions -98, -129, -184 and -541 in combination and a selected number of alternative gRNA2.1 combinations (dCas9:VPR-Ms2:VPR and dCas9:TV-Ms2:VPR). As shown in Figure 4C, the results confirm that dCasEV2.1 achieved the best activation rates also when using the *pSIMTB* as target promoter.

Next, the activation of the *SIDFR* promoter by dCas9-PTAs was analysed. The *SIDFR* gene is involved in the flavonoid biosynthesis pathway, and it is strongly induced by natural transcription factors of the MYB family. Three gRNAs targeting the positions -376, -300 and -150 were generated and tested with scRNA-2.1. The *pSIDFR* activation assay was also interrogated with *SIANT1*, a MYB factor that naturally activates the *SIDFR* gene in tomato. Best activation rates were obtained with gRNA that targets at position -150 (x100 fold) (Figure 4E), reaching transcriptional activation values very similar to that

observed with the induction of *SIANT1*. As in previous promoters assayed, dCasEV2.1 outperformed other AD combinations also in the case of p*SIDFR* (Figure 4F).

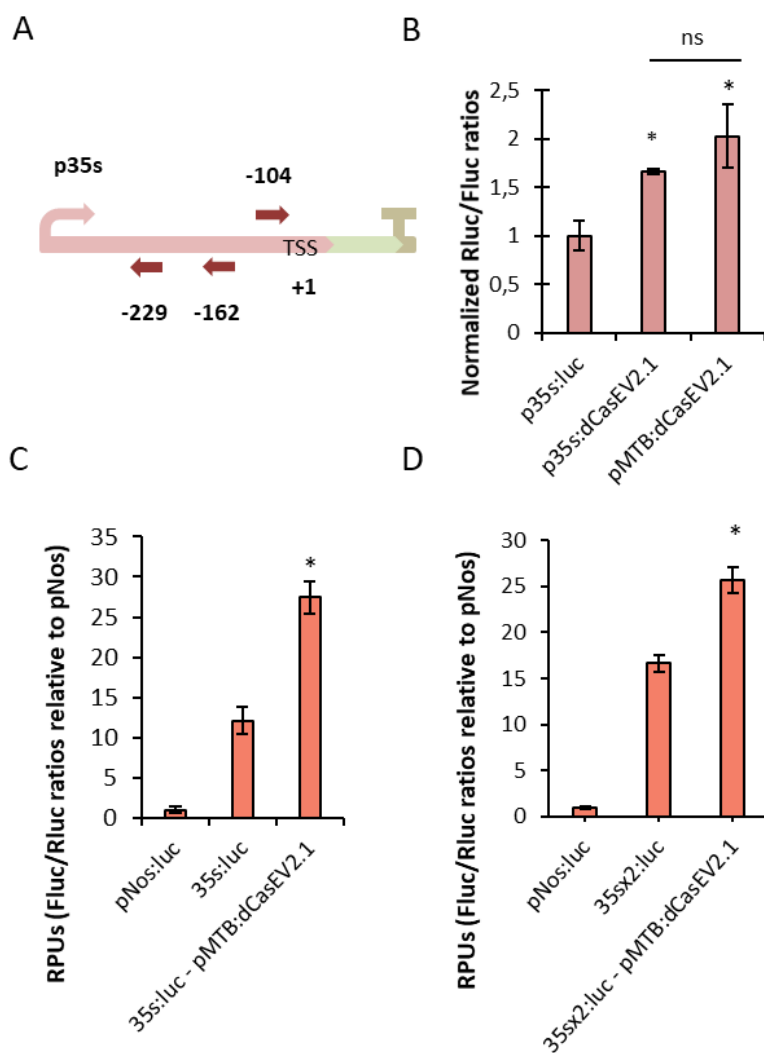


**Figure 4: Transient activation of *SIMTB* and *SIDFR* promoter by dCas9-PTAs. A)** Representation of the reporter p*SIMTB* with gRNAs targeted at positions -98, -129, -184

and -541 with dCasEV2.1 B) RTAs obtained with gRNAs targeting reporter *pSIMTB* at positions -98, -129, -184, and -541, individually and in combination, using dCasEV2.1. C) RTAs obtained upon activation with the best combinations of TADs (dCas9:EDLL-MS2:VPR, dCas9:VPR-MS2:VPR and dCas9:TVMS2:VPR) using scRNA-gRNA2.1 strategy and targeting the reporter *pSIMTB* at positions -98, -129, -184, and -541. The Student's test was performed for comparing the Rluc/Fluc values of the reference Mtb:Luc sample with the Rluc/Fluc values obtained from the activated Mtb:luc samples. Asterisks indicate Student's-test significant values ( $p < 0.05$ ). D) Representation of the reporter *pSIDFR* with gRNAs targeted at positions -150, -300, and -376. E) Relative transcriptional activities (RTAs) were obtained with gRNAs targeting reporter *pSIDFR* at positions -150, -300, and -376 individually, using dCasEV2.1. In this experiment, MYB TF *SIANT1* was included to test the RTA obtained with the natural transcription factor of *pSIDFR*. F) RTAs obtained upon activation with the best combinations of TADs (dCas9:EDLL-MS2:VPR, dCas9:VPR-MS2:VPR and dCas9:TV-MS2:VPR) using scRNA-gRNA2.1 strategy and targeting the reporter *pSIDFR* at position -150. The Student's test was performed for comparing the Rluc/Fluc values of the reference *pSIDFR*:Luc sample with the Rluc/Fluc values obtained from the activated *pSIDFR*:Luc samples. Asterisks indicate Student's-test significant values ( $p < 0.05$ ). RTAs are measured as relative promoter units (RPU) calculated as the Fluc/Rluc ratios for each construct normalized with the Fluc/Rluc ratio of a pNos:Luc reporter transformed in parallel. The pNos:Luc measure is a reference sample with a dCasEV2.1 and a non-specific gRNA. Bars represent average RTAs  $\pm$  SD, n=3

The range of activations obtained with the dCasEV2.1 tool on different promoters expands the applications that could be generated in plants. For example, the results obtained when activating promoters with a high basal activity suggest that this system could be used to improve the recombinant protein production in plants. Although the results with the *pSIMTB* promoter are very promising due to the high levels of activation achieved, the CaMV35S promoter was chosen as a promising target for activation due to it being one of the most widely used strong constitutive promoters. However, before testing directly the dCasEV2.1 tool over the p35S:Luc employing the same Fluc/Rluc reporter methodology, a few considerations had to be taken. First, the normalizer renilla reporter is driven under the control of the CaMV35S promoter, so the reference p35S:Luciferase-p35S:Renilla reporter is not suitable for being activated. Second, the dCasEV2.1 tool components, dCas9 and MS2, are also driven under the control of the

CaMV35S promoter, so their activation can direct the dCasEV2.1 action over its own transcriptional unit. To solve this, a new version of the dCasEV2.1 driven by the p*SIMTB* promoter was developed, due to this promoter having shown high basal activity in *N. benthamiana*. In parallel, three gRNA targeting the positions -104, -162 and -229 of the CaMV35S promoter were designed (Figure 5A) and assembled through the multiplexing 2.1 strategy. As a first comparison, the two versions of dCasEV2.1, pMTB:dCasEV2.1 and p35S:dCasEV2.1, were tested over the Fluc/Rluc reporter, but in this case, interrogating the renilla activity, which is driven by CaMV35S promoter, and employing the luciferase driven by *Nos* promoter as the normalizer. As shown the Figure Figure 5B, both versions of dCasEV were able to increase the transcriptional rates of the renilla reporter, with no statistical significance between them. Finally, in order to make standard the measurements, new Fluc/Rluc reporters were developed where the renilla is driven by the *Nos* promoter. This approach was developed to represent normalized Fluc/Rluc values as standardized values (RPU). Also, a luciferase reporter driven by a double *CaMV35* promoter (35Sx2) was included to check if the level of activation achieved can be further increased. The results showed significant activation of 35S:Luciferase (Figure 5C) and 35Sx2:Luciferase (Figure 5D), however, the luciferase reporter driven by double CaMV35S promoter does not show an increased activation compared with the activation levels that the p35S:Luciferase reporter can reach. These results suggest that these activation rates are the maximum that the CaMV35S promoter could offer in *N. benthamiana*.



**Figure 5: Transient activation of CaMV35 promoter by dCas9EV2.1.** A) Representation of the reporter pCaMV35 with gRNAs targeted at positions -104, -162 and -229 with dCasEV2.1. B) Rluc/Fluc ratios normalized to pNos:Ren obtained upon activation with dCasEV2.1 driven by either the CaMV35s or the *SIMTB* promoters and targeting positions -104, -162 and -229 of the CaMV35S promoter. The p35S:Luc measure is a reference sample with a dCasEV2.1 and a non-specific gRNA. C) RTAs obtained with gRNAs targeting reporter p35S:Luc at positions -104, -162 and -229 with dCasEV2.1. D) RTAs obtained with gRNAs targeting reporter 2xp35S:Luc at positions -104, -162 and -229 with dCasEV2.1. RTAs are measured as relative promoter units (RPUs) calculated as the Fluc/Rluc ratios for each construct normalized with the Fluc/Rluc ratio of a pNos:Luc reporter transformed in parallel. The pNos:Luc measure is a reference sample with a dCasEV2.1 and a non-specific gRNA. Asterisks indicate Student's-t-test significant values ( $p < 0.05$ ). Bars represent average RTAs  $\pm$  SD,  $n=3$ .



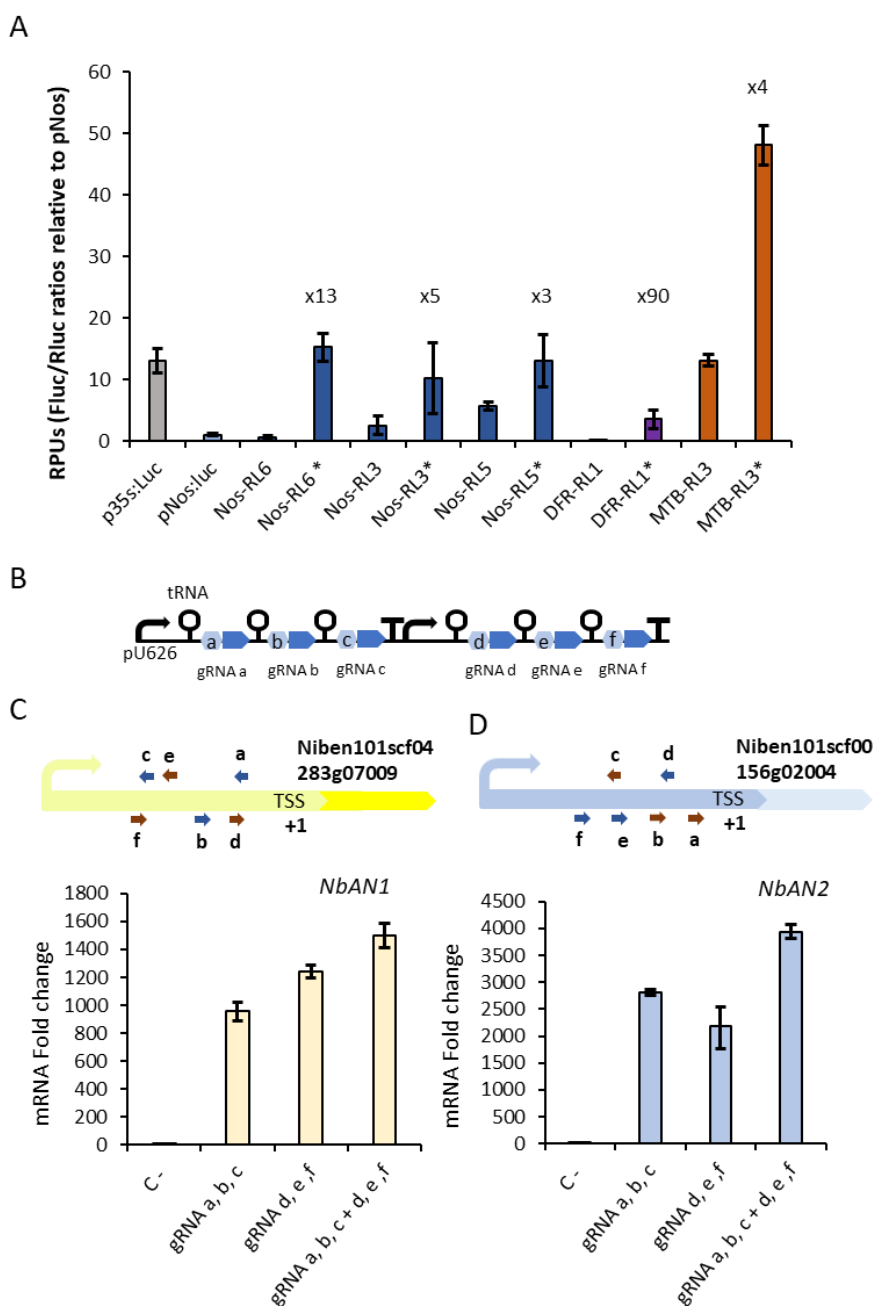
## **Transcriptional activation of endogenous genes in *Nicotiana benthamiana***

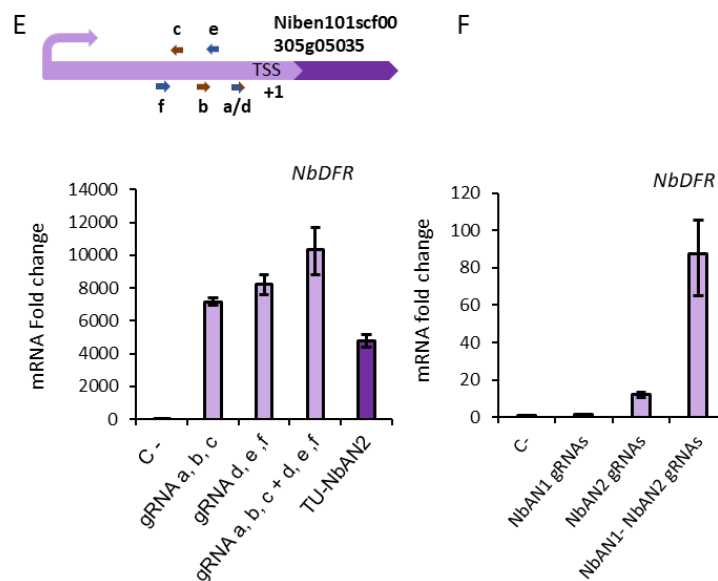
Transient experiments with *Agrobacterium*-delivered T-DNAs allow fast combinatorial dCas9-PTA transactivation assays. However, the episomal status of at least part of the transcriptionally active T-DNAs used as reporters may lead to regulatory dynamics different from those of stably integrated genes, which are often influenced by nucleosome positions and subjected to epigenetic regulation. Therefore, we next studied dCasEV2.1 activity on chromatin-integrated genes, using two types of strategies: (i) by targeting reporter promoters integrated into transgenic lines and (ii) by assessing activation of endogenous *N. benthamiana* genes.

Transgenic reporter lines were generated by transforming *N. benthamiana* with T-DNA constructs used in transient experiments plus the addition of a *KanR* module. Single-copy T2 lines Nos-RL3, Nos-RL5 and Nos-RL6 (for *Nos* promoter), MTB-RL3 (for *pSIMTB* promoter) and DfrRL1 (for *pSIDFR* promoter) were selected for activation tests after analysing their *KanR* segregation. For each line, leaves were agroinfiltrated using the dCasEV2.1 activating construct with or without target-specific gRNAs, and the activation levels were measured as Fluc/Rluc ratios. As shown in Figure 6A, activation rates were in line with what was observed in transient experiments. Chromatin-integrated *pNos* was induced 3-13 fold reaching maximum levels similar to that of the CaMV35S promoter; the *pSIMTB* line, showing higher basal levels, was only activated 3X but reached record expression levels in absolute terms (as compared with an agroinfiltrated CaMV35S promoter used as reference). DFR-RL1 line having the lowest basal luminescence rates showed the highest activation rate, up to 90 fold.

Next, we analysed the ability of dCasEV2.1 to induce endogenous *N. benthamiana* genes. For this purpose, we targeted the endogenous *N. benthamiana* orthologue of the *SIDFR* gene, *NbDFR*, and two transcription factors involved in polyphenol biosynthesis (the MYB factor *NbAN2* and the bHLH protein *NbAN1*), which jointly activate *NbDFR* expression (Spelt *et al.*, 2000). In these experiments we followed a multiplexing strategy,

targeting each gene with groups of three gRNAs expressed in multicistronic transcripts as depicted in Figure 6B. Results show a strong 10000 fold activation of *NbDFR* with dCasEV2.1 loaded with six gRNA combinations, outperforming that obtained with endogenous MYB *NbAN2* (Figure 6E). Similarly, *NbAN1* and *NbAN2* reached maximum (>1000 fold) activation rates with a 6X multiplex gRNA combination (Figure 6C, 6D). Simultaneous targeting of *NbAN1* and *NbAN2* led to an 80X secondary activation of *NbDFR* as shown in Figure 6F.





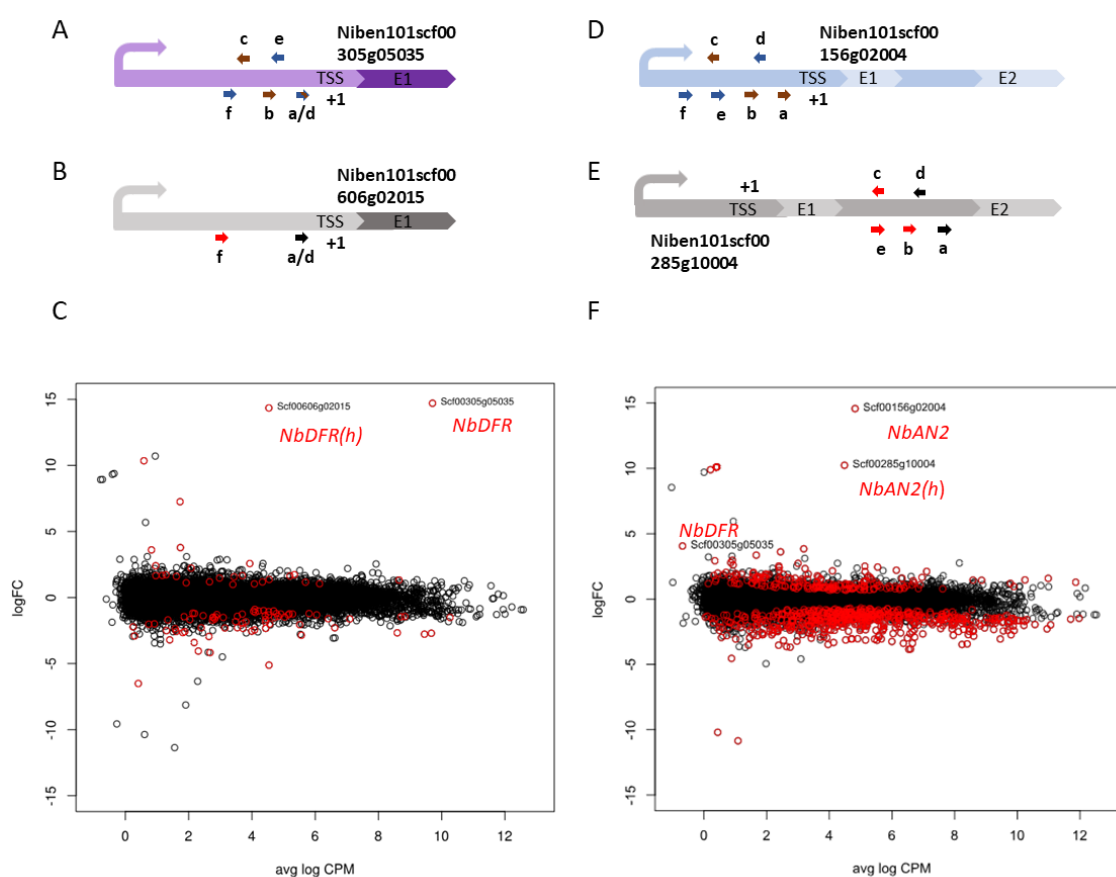
**Figure 6: Transient activation of transgenic reporter lines and endogenous genes in *N. benthamiana* with dCasEV2.1** A) Relative transcriptional activities (RTAs) obtained for the activation of *Nos*, *SIDFR* and *SIMTB* reporter transgenic lines using dCasEV2.1. The *Nos*-RL6, *Nos*-RL3 and *Nos*-RL5 were activated targeting the *Nos* promoter at positions -161 and -211. *DFR*-RL1 was activated targeting the *SIDFR* promoter at position -150. *MTB*-RL3 was activated targeting the *SIMTB* promoter at positions -98, -129, -184 and -541. Asterisks represent the reporter line activation using dCasEV2.1 RTAs are measured as relative promoter units (RPUs) calculated as the *Fluc*/*Rluc* ratios for each construct normalized with the *Fluc*/*Rluc* ratio of a *pNos:Luc* reporter transformed in parallel. The *pNos:Luc* measure is a reference sample with a dCasEV2.1 and a non-specific gRNA. Bars represent average RTAs  $\pm$  SD, n=3. B) Representation of the multiplexing gRNA strategy structure employed to express 6 gRNA2.1. C) mRNA fold change obtained targeting the endogenous gene *NbAN1* with dCasEV2.1 through the multiplexing gRNA strategy after 4 dpi. Represented gRNAs target the promoter of *NbAN1* in the positions: a (-101), b (-173), c (-242), d (-120), e (-219) and f (-242). D) mRNA fold change obtained targeting the endogenous gene *NbAN2* with dCasEV2.1 through the multiplexing gRNA strategy after 4 dpi. Represented gRNAs target the promoter of *NbAN2* in the positions: a (-103), b (-175), c (-196), d (-145), e (-198) and f (-252). E) mRNA fold change obtained targeting the endogenous gene *NbDFR* with dCasEV2.1 through the multiplexing gRNA strategy after 4 dpi. Represented gRNAs target the promoter of *NbDFR* at the positions: (-88), b (-125), c (-217), e (-198) and f (-248). In this experiment, a MYB TF *NbAN2* driven by the *CaMV35s* promoter was included to test the RTA obtained with the natural transcription factor targeting *NbDFR*. (F) mRNA fold change obtained by the indirect activation of *NbDFR* at 7 dpi, by targeting *NbAN1* and *NbAN2*

promoter genes with gRNAs a, b, c and d, e, f using dCasEV2.1. Bars represent average mRNA fold change  $\pm$  SD, n=3.

The strong activation observed for endogenous genes when targeted with dCasEV2.1 prompted us to investigate if enhanced transcriptional activation had occurred at the expense of specificity, therefore triggering non-specific or global changes in the transcriptome besides those on the designated target. To investigate this, we performed RNAseq analysis of leaf samples treated with dCasEV2.1 targeting the *NbDFR* and *NbAN2* promoters. In parallel, leaf samples were treated with dCasEV2.1 but no gRNAs were used as a control. The *NbDFR*-induced samples target *NbDFR* (Niben101Scf00305g05035) promoter with 6xgRNA (Figure 7A), following the gRNA design used in previous assays. Also, a search for potential off-targets was performed in order to find other possible genes activated in the transcriptome (Supplementary Table 8). A second gene (Niben101Scf00606g02015), labelled as *NbDFR*(h), known to be a closely related paralogue of selected *NbDFR*, shares a 100% identical gRNA target sequence in its promoter and another potential target sequence with two mismatches (Figure 7B). The differential gene analysis is represented in Figure 7C and plots on the y axis the fold change between *NbDFR*-induced and control samples, whereas the x-axis shows the log CPM (counts per million) indicative of absolute expression levels for each gene. As can be observed, the transcriptome of the dCasEV2.1/*NbDFR* sample, compared to the control, remains virtually unchanged except for the *NbDFR* gene itself and a second gene, *NbDFR*(h). Strikingly, *NbDFR* is induced from extremely low expression levels to the top 50 mRNA in the transcriptome. No significant category enrichments of upregulated genes and downregulated genes (Supplementary Table 8) were found in *NbDFR*-activated samples. The *NbAN2*-induced samples target the *NbAN2* (Niben101Scf00156g02004) promoter with 6 gRNAs (Figure 7D). A search for potential off-targets derived from these gRNAs was also performed (Supplementary Table 7). A *NbAN2* homologue (Niben101Scf00285g10004), labelled as *NbAN2*(h), also shows two 100% identical gRNA target sequences and three homologous sequences with at least 3 mismatches (Figure 7E). The differential gene analysis plotted in Figure 7F shows the equivalent representation obtained with dCasEV2.1 activation of the *NbAN2* gene. *NbAN2* is now detected as the most activated gene in the genome. In this case, a second

gene with a smaller fold change was observed, corresponding to *NbAN2(h)*. Significant secondary activation of *NbDFR* was also detected at the transcriptome level, with much lower log CPM and fold change values.

As expected for a TF, *NbAN2* activation was accompanied by more significant changes in the transcriptome, although most of them had fold changes far below those observed in the targeted genes. GO analysis showed enrichment in defence categories in upregulated genes, and primary metabolism and photosynthesis-related categories in downregulated genes (Supplementary Table 9).



**Figure 7. Differential gene expression analysis in samples activated with dCasEV2.1** A) Representation of the *NbDFR* (Niben101scf00305g05035) promoter and exon 1 region with gRNAs targeted at positions: a/d (-88), b (-125), c (-217), e (-198) and f (-248) to TSS. (B) Representation of homologous *NbDFR(h)* (Niben101scf00606g02015) gene targeted at positions a/d (-89) and f (-261) to TSS. Black coloured gRNAs have 100% identity in the homologous gene sequence, red coloured are off-target gRNAs with at least 2 mismatches and an off-target score <5. C) Differential gene expression plot between the dCasEV2.1–*NbDFR* condition and control condition. The logFC axis represents the log of the fold change

of the gene expression obtained in the samples that target the *NbDFR* with dCasEV2.1. Positive logFC values represent gene induction and negative logFC values represent gene repression. Avg logCPM is a normalized value that represents the gene abundance in the transcriptome. D) Representation of the *NbAN2* (Niben101scf00156g02004) promoter and exon 1 region with gRNAs targeted at positions: a (-103), b (-175), c (-196), d (-145), e (-198) and f: (-252) to TSS. E) Representation of homologous *NbAN2* (Niben101scf00285g10004) gene targeted at positions: a (+496), b (+424), c (+403), d (+474) and e (+401) to TSS. Black coloured arrows represent gRNAs with 100% identity in the homologous gene sequence, red coloured arrows represent gRNAs off-targets with at least 3 mismatches and off-target score <2. D) Differential gene expression plot between the dCasEV2.1–*NbAN2* condition and control condition. The logFC axis represents the log of the fold change of the gene expression obtained in samples that target the *NbAN2* with dCasEV2.1. Positive logFC values represent gene induction and negative logFC values represent gene repression. Avg logCPM is a normalized value that represents the gene abundance in the transcriptome.

## DISCUSSION

Ideally, programmable transcriptional activators should be potent, orthogonal, and able to activate a wide spectrum of targets. Maximum potency is an obvious engineering objective in the first place, as lower activation rates can be later obtained if needed by simply modifying the optimal parameters. The first generation of dCas9-PTAs tested in plants showed limited activation potency. Based on single translational fusions of dCas9 with activation domains, they have been soon surpassed by more sophisticated architectures involving tandem repeats of TAD and showing improved activation rates. Li *et al.* (2017a) reported potent activation using the so-called TV combination, which comprised a VP128 tandem repeat with up to six copies of the TALE TAD motif (dCasTV strategy). This TV autonomous activator arranged in a SAM strategy (linked to the gRNA using the MS2 aptamer) produced strong activation rates in the *OsGW7-Luc* reporter construct assayed in rice protoplasts. Similarly, Lowder *et al.* (2018a) reported strong activation results with the gRNA2.0 strategy, with maximum activation levels obtained with the so-called CRISPR-Act2.0 approach, which comprises dCas9:VP64 coupled with MS2:VP64 via gRNA2.0 system.

In a search for further improvements in the performance of dCas9-PTAs, we designed, to our knowledge, the most comprehensive analysis of dCas9-PTAs tested in plants, comprising a total of 43 combinations of TADs displayed in different protein architectures, including those showing the strongest activation in previous studies. When assayed against pNos:Luc reporter, the selected dCasEV2.1 PTA almost tripled the activation levels obtained with dCas9:TV and CRISPR-Act2.0 PTAs. Furthermore, a record activation of four orders of magnitude was reached when dCasEV2.1 targeted the endogenous *NbDFR* gene. The clue for this strong activation capacity seems to be an additive effect of the TADs employed (EDLL plus VPR) and the favourable gRNA2.1 loop. EDLL/VPR pairing was tested in the two possible combinations (dCas9:EDLL-MS2:VPR and vice versa), and in both of them, the EDLL/VPR pair yielded the highest activation rates as measured with pNos:Luc. On the other hand, the modification introduced by serendipity in the gRNA2.1 loop consistently yielded activation rates between 20-35% higher than the original gRNA2.0 loop in all different combinations assayed. The causes of this effect are unknown, although we speculate that the modification could favour the stability of the gRNA scaffold in the plant cell. Besides dCasEV2.1, the remaining TAD combinations analysed here will be useful, alone or in combination with less efficient/distant gRNAs, to achieve intermediate activation levels when required. All the DNA elements shown here are integrated into the GoldenBraid 3.0 (GB3.0) modular cloning system and conform with the phytobrick standard (Patron *et al.*, 2015; Vazquez-vilar *et al.*, 2017). GB3.0 enables fast combinatorial assembly, easy multigene engineering, exchange and reuse of standard parts and straightforward multiplexing of Cas9 gRNAs.

dCasEV2.1 successfully activated all promoter targets assayed, either transiently expressed or stably integrated into the chromosome driving expression of reporter constructs or endogenous genes, suggesting a wide activation spectrum that will need to be confirmed in further experiments involving a larger number of targets. However, the activation rates differed strongly among promoters. By referring all Fluc/Rluc measurements to those obtained with a standard pNos:Luc reporter phytobrick assayed in parallel, it was possible to obtain a general and comparative view of the transcriptional levels conferred to each of the three promoters assayed in transient

---

expression analysis. A relationship between the basal expression levels of the promoter and the induction rates obtained with dCasEV2.1 became apparent by analysing these results, as has been suggested earlier (Haran *et al.*, 2000; Ede *et al.*, 2016). Thus, although the strongest induction rate was obtained with p*SIDFR*, this promoter could not reach levels above 5.9 RPU in its fully activated state, which is half the activity measured for the CaMV35S-derived phytobrick. In contrast, p*SIMTB*, showing much more modest activation rates (8X), achieved instead remarkable RTA levels (80 RPU), six times stronger than the 35S phytobrick. This was also confirmed in the stably transformed reporter lines, with p*SIMTB* showing extremely high Fluc/Rluc levels upon induction, although in this case standard measurements are not possible due to the lack of a reference free of positional effects. The record Fluc/Rluc expression levels obtained with dCasEV2.1-activated p*SIMTB* suggest that this strategy could be exploited to boost yields of recombinant proteins and/or metabolites in biofactory and/or metabolic engineering approaches.

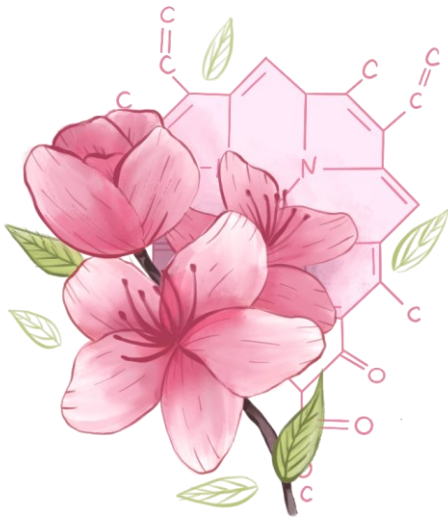
From an engineering standpoint, one of the most important features of PTAs is the ability to combine strong activation with genome-wide specificity. Off-target activities have been often regarded as the Achilles heel of Cas9 strategies, but their effect in plant PTA-regulation has not been established. Li *et al.* (2017a) showed almost no influence of dCas9:TV on the transcriptome profile of *Arabidopsis* protoplasts; however, in that study, the activation rates reached by the target gene were rather modest (below 10 fold). With RT-PCR data of *NbDFR* and *NbAN* genes showing activation rates between 1000-10000 fold, it was important to determine the effect on specificity. Deliberately, we selected *NbDFR* and *NbAN2* target genes as representative of two distinct categories of potential actuators: (i) enzyme-coding genes (*NbDFR*) as final actuators with no transcriptional regulatory roles known, and (ii) transcriptional factors (*NbAN2*) representing a regulatory node with connections in the transcriptome. The results obtained with *NbDFR*-treated transcriptome demonstrate that strong activation with dCasEV2.1 is not incompatible with genome-wide specificity, as an almost invariant gene expression profile was observed when compared with a control transcriptome, with the exception of an *NbDFR* homologue that served as additional proof of specificity. In contrast, dCasEV2.1-*NbAN2* treatment resulted in wider transcriptome changes,



generally consisting of low but significant changes in expression ratios. The changes observed in dCasEV2.1-*NbAN2* are interpreted as a result of the *NbAN2* regulatory role. This is confirmed by the presence of *NbDFR* among the pool of genes significantly induced in dCasEV2.1-*NbAN2* treatment, although with a fold change (and absolute expression levels) much lower than those observed in direct dCasEV2.1-*NbDFR* treatment. Interesting, dCasEV2.1-*NbAN2* downregulated genes are enriched in GOs related to primary metabolism and photosynthesis, whereas upregulated genes are enriched in categories related to plant defence. It was earlier described that agroinfiltration of *N. benthamiana* Lab strain with Rosea-like MYB factors fails to engage a fully active anthocyanin biosynthesis pathway and initiates instead a necrosis-like programmed cell death process with activation of defence pathways, a reaction fully compatible with the observed decrease in the expression of genes involved in primary metabolism (Outchkourov *et al.*, 2014).

In sum, we show here an improved PTA toolset that combines maximum potency with genome-proven orthogonality. This toolset has been shown to act efficiently at two different levels: with a heavily connected master regulator and with a final actuator enzyme. Both examples have strong potential applied implications. Targeting master regulators can be used to couple agronomically-relevant regulatory networks (e.g. defence, phase transition) to new external inputs (e.g. agrochemicals). On the other hand, custom-activation of individual enzymatic steps will facilitate steering metabolic flows towards selected final products.





## Chapter 2

# Custom-made design of metabolite composition in *N. benthamiana* leaves using CRISPR activators

**Selma, S.**, Sanmartín, N., Espinosa-Ruiz, A., Gianoglio, S., Lopez-Gresa, M., Vázquez-Vilar, M., Flors, V., Granell, A. and Orzaez, D. (2022), Custom-made design of metabolite composition in *N. benthamiana* leaves using CRISPR activators.

Plant Biotechnology Journal (2022), DOI: 10.1111/pbi.13834

My contribution to this work was essential to this publication. I contributed to most of the analysis performed and a major part of the manuscript writing.







## ABSTRACT

Transcriptional regulators based on CRISPR architecture expand our ability to reprogram endogenous gene expression in plants. One of their potential applications is the customization of plant metabolome through the activation of selected enzymes in a given metabolic pathway. Using the previously described multiplexable CRISPR activator dCasEV2.1, we assayed the selective enrichment in *Nicotiana benthamiana* leaves of four different flavonoids, namely naringenin, eriodictyol, kaempferol, and quercetin. After careful selection of target genes and guide RNAs combinations, we created successful activation programs for each of the four metabolites, each program activating between three and seven genes, and with individual gene activation levels ranging from 4- to 1500-fold. Metabolic analysis of the flavonoid profiles of each multigene activation program showed a sharp and selective enrichment of the intended metabolites and their glycosylated derivatives. Remarkably, principal component analysis of untargeted metabolic profiles clearly separated samples according to their activation treatment, and hierarchical clustering separated the samples into five groups, corresponding to the expected four highly enriched metabolite groups, plus an un-activated control. These results demonstrate that dCasEV2.1 is a powerful tool for re-routing metabolic fluxes towards the accumulation of metabolites of interest, opening the door for the custom-made design of metabolic contents in plants.

**Keywords:** CRISPR, Metabolic Engineering, *Nicotiana benthamiana*, Flavonoid Pathway.

## INTRODUCTION

Traditionally, plant breeding has contributed to the generation of crop varieties adapted to changing external conditions and to consumers' demands by selecting favourable genetic variants in a species' genomic pool, or by introducing new genetic traits through transgenesis or mutation by CRISPR (Jisha *et al.*, 2015; Maioli *et al.*, 2020). In recent years, the need for new adaptations has grown exponentially, fostered by climate change and human population dynamics, and the pressure to explore new breeding strategies has increased (Godwin *et al.*, 2019). In this context, new breeding concepts inspired in Synthetic Biology propose the development of new programmable traits where physiological outputs (e.g., a developmental phase transition, or the activation of a metabolic pathway) occur as a response to endogenous or external inputs (e.g., a chemical cue, or an electromagnetic signal) perceived by synthetic sensors, and operated by engineered genetic operators (e.g., logic gates, toggle switches, oscillators, etc.) (McKenzie *et al.*, 1998; Ochoa-Fernandez *et al.*, 2016; Bernabé-Orts *et al.*, 2020). To produce the desired physiological output, operators need to be transcriptionally connected to a selected group of final actuators (e.g., a group of enzymes) that ultimately generate the designed phenotypic changes. Natural gene circuits have evolved intricate regulatory cascades of transcription factors (TFs) that connect operators with downstream actuators in a concerted manner, jointly generating a consistent physiological response. Among the many challenges facing Plant Synthetic Biology, a key one is to acquire the ability to customize the connections between synthetic operators and endogenous actuators in ways that are different from those designed by evolution but convenient for agriculture. Examples of new "synthetic" connections are the modification of flowering time (Papikian *et al.*, 2019), the activation of an anticipated response to a forecasted biotic/abiotic stress (Chen *et al.*, 2020), or the customization of metabolic composition (Llorente *et al.*, 2020). Natural transcription factors, which are often used as connection hubs in traditional genetic engineering approaches (Xie *et al.*, 2006; Naing *et al.*, 2017), have a limited capacity for circuit customization due to their hardwired DNA binding specificities, which impede free selection of the downstream genes to be regulated. Recently, a new type of



programmable transcriptional regulators (PTRs) based on CRISPR/Cas has emerged that allows easy customization of both DNA binding and transcriptional regulatory activities. The nuclease-inactivated CRISPR/Cas9 (dCas9) (Maeder *et al.*, 2013) architecture enables the combination of autonomous transcriptional activation domains (TADs) with the DNA-binding specificities of Cas9, which can be programmed through a small guide RNA with minimum engineering efforts. In plants, several strategies to build these potent Programmable Transcriptional Activators (PTA) based on dCas9 have been described (Park *et al.*, 2017; Lowder *et al.*, 2018a; Li *et al.*, 2019; Pan *et al.*, 2021b) that trigger efficient activation of target endogenous genes. The two main advantages of dCas9-PTA are (i) their high accuracy, reaching single-gene specificity levels as reported recently for a transcriptomic analysis showing negligible off-target activation (Li *et al.*, 2017a); and (ii) their amenability for multiplexing (Vazquez-Vilar *et al.*, 2016; Lowder *et al.*, 2017). This latter feature enables the concerted activation of multiple actuators simultaneously by simply loading the cell with several gRNAs, each one targeting a different promoter or a different position within a given promoter. The practical implications of multiplexing PTRs are widespread, from the design of synthetic regulatory cascades to the re-routing of endogenous metabolic fluxes. However, the capacities of PTRs in producing new-to-nature phenotypes are just starting to be explored, and no examples exist yet where PTRs are applied to re-route biosynthetic fluxes. Mastering the regulation of metabolic pathways would open the way for the customization of plant composition, a possibility with many implications in food and feed design, as well as in the development of plant biofactories.

The phenylpropanoid pathway is an alluring bioengineering target due to its pharmaceutical and industrial interest (Neelam *et al.*, 2020). Besides, it is a highly branched pathway in plant secondary metabolism, offering interesting opportunities for biotechnological regulation and fit to test new technological approaches. The pathway can be divided into different parts (Figure 1). The “general” section of the pathway generates cinnamic acid, coumaric acid, and 4-coumaroyl-CoA, the basic backbone products derived from phenylalanine, thus providing the core structures for the biosynthesis of all flavonoids as well as for other phenylpropanoid branches like the lignin pathway (Nabavi *et al.*, 2020). The second group of reactions leads to the

---

condensation and subsequent cyclization of the core structure, generating the first flavanones of the pathway. Finally, flavanones serve as substrates for multiple reactions that form the distinct branches of the flavonoid pathway, including those making flavonols, anthocyanins, isoflavonoids, etc. (Nabavi *et al.*, 2020). In plants, the main flavanones are naringenin, eriodictyol, and hesperetin, which can also be found in their glycosylated forms and whose distribution changes between species and tissues. For example, in grapefruit and tomato, naringenin glycosyl-conjugated compounds are the predominant flavanones present in the fruits, while in some species of citrus, like mandarin or lime, hesperetin glycosyl-conjugates are the most abundant flavanones (Khan *et al.*, 2014). This variation is due to the different combinations of gene expression patterns in the upstream part of the pathway. Similarly, the differential expression of downstream enzymes governs the predominant accumulation of flavonols (e.g., kaempferol, quercetin) or anthocyanins (e.g., delphinidin, pelargonidin), shaping important traits such as fruit colour, antioxidant activity, etc.

The flavonoid pathway has been the subject of remarkable metabolic engineering interventions mainly by making use of native transcription factors (Dias and Grotewold, 2003; Park *et al.*, 2021). The enzymes in the pathway are frequently regulated by a triad of TFs comprising an MYB TF, a bHLH, and a WD repeat component (Zhao *et al.*, 2013). In many cases, the overexpression of the MYB factor alone or in combination with a bHLH is sufficient to ectopically activate an entire branch of the pathway (Liu *et al.*, 2015). MYB factors show a certain degree of specificity for activating the biosynthesis of different flavonoid subgroups. Thus, whereas the ROSEA transcription factor activates enzymes in the anthocyanin branch in tomato and tobacco (Fresquet-Corrales *et al.*, 2017; Vu and Lee, 2019), the *Arabidopsis thaliana* MYB12 factor activates the enzymes of the flavonol subgroup, leading to the accumulation of kaempferol and rutin (quercetin glycosylate) as main products (Misra *et al.*, 2010; Zhang *et al.*, 2015a). With the elucidation of the specificities of natural TFs from different species, followed by their recombinant expression, Butelli and co-workers obtained multi-level engineering of flavonoid compounds in tomato (Butelli *et al.*, 2008). As shown in their work, the engineering precision obtained with native TFs reaches its limit at the “subgroup” level, as native TFs collectively activate those endogenous genes sharing similar regulatory

sites in their 5' untranscribed regions, and these genes usually correspond to enzymes in the same branch of the pathway (e.g., flavanols, anthocyanins or flavanones branches). To achieve higher precision levels in plant metabolic engineering, up to the level of individual enzymes (and metabolites), endogenous TFs seem not to be fit for the purpose, and it would be necessary to break the evolutionary constraints and incorporate the type of single-gene specificity offered by PTAs.

In this work, we aimed to explore the efficiency and the precision limits of dCas9-PTAs for engineering the specialized metabolism in plants, using the so-called dCasEV2.1 programmable activator previously developed in our group (Selma *et al.*, 2019). This PTA is based on the CRISPR-scRNA strategy (Koneremann *et al.*, 2015a; Zalatan *et al.*, 2015) and comprises two gene modules, a constant module with two constitutively expressed transcriptional units (TUs), and a variable module carrying the gene-specific activation program. In the constant module, the first TU produces a deactivated Cas9 translationally fused to an EDLL domain, a plant activation domain isolated from the ethylene response factor (Tiwari *et al.*, 2012). The second TU expresses the VPR activation domain (Chavez *et al.*, 2015), a tandem fusion of the viral domains VP64, P65 and RTA, fused to the MS2 phage RNA aptamer-binding protein.

Initially, we first selected individual activation programs (i.e., single polycistronic constructs expressing up to six gRNAs) for ten different target enzymes distributed in the general flavonoid pathway and the flavanone/ flavonol branch of the pathway. Then, we combined those enzymes in four groups, each group leading to the biosynthesis of a different flavonoid compound as the final product. Four multigene activation programs (i.e., combinations of polycistronic gRNA constructs targeting several genes simultaneously) were constructed and assayed transiently in *N. benthamiana*, each program designed to specifically activate the genes in one of the four groups. As a result, four different and highly specific metabolic profiles were generated in the leaf, with a highly predominant accumulation of the expected target products in each of the assayed combinations. These results show that dCasEV2.1 raises metabolic engineering to a new precision level, opening the door for true customization of plant metabolic composition.

## MATERIALS AND METHODS

### gRNA design

The selection of the protospacer sequences for the gRNA design was performed following the previously described criteria (Selma *et al.*, 2019), where the *N. benthamiana* genes were targeted with one or two sets of three polycistronic gRNAs each. Once the position of the TSS was determined for each gene, the protospacer sequences were selected within the activation window located between -100 and -300 bp upstream of the TSS. The selection was performed taking into consideration the best on-target score and the low off-target score for each gRNA employing the Benchling software tool for CRISPR design ([www.benchling.com](http://www.benchling.com)). The protospacer sequences selected for each gene presented a minimum distance of 50 pb between them for avoiding the overlapping of the binding area occupied by dCas9 complexes and for covering the maximum length of the activation window. All the protospacer sequences employed in this work are listed in Supplementary Table 10.

### DNA constructs

All plasmids used in this work were assembled using GoldenBraid (GB) cloning (Sarrion-Perdigones *et al.*, 2013). The DNA sequences of the constructs generated in this work are available at [www.gbcloning.upv.es](http://www.gbcloning.upv.es) by entering the IDs provided in Supplementary Table 11. Briefly, the multiplexing vectors used for this work were generated as GB level -1 vectors and previously described in Selma *et al.* (2019). The level -1 vectors, pVD1\_M1-3pTRNA scf 2.1 (GB1436) pVD1\_M2-3pTRNA scf 2.1 (GB1437), and pVD1\_M3-3pTRNA scf 2.1 (GB1438) were used to assemble the protospacer sequences occupying the first, second, and third positions in the final gRNA assembly. For GB gRNA assemblies, two partially complementary primers containing the protospacer sequence were designed at [www.gbcloning.upv.es/do/crispr/multi\\_cas9\\_gRNA\\_domesticator\\_1](http://www.gbcloning.upv.es/do/crispr/multi_cas9_gRNA_domesticator_1). Primers were resuspended in water to final concentrations of 1  $\mu$ M and equal volumes of forward and reverse primer were mixed. The mixture was incubated at room temperature for 5 min for the hybridization of the primer pair. 1  $\mu$ l of the primer mix

was included in a BsmBI restriction–ligation reaction with 75 ng of pUPD2 and 75 ng of the corresponding level –1 vector for the assembly of tRNA-protospacer-scaffold units in level 0 plasmid. The next step consists of the assembly of the multiplexing gRNA expression cassette in a level 1 vector. For level 1 assemblies, 75 ng of level 0 gRNA for each position, 75 ng of U626 promoter (GB1001), and 75 ng of pDGB3 $\alpha$  destination vector were included in a BsaI restriction-ligation reaction.

The combination of level 1 multiplexed gRNAs and dCasEV2.1 TUs were performed by binary BsaI or BsmBI restriction–ligation reactions to obtain all the level  $\geq 1$  assemblies as described in Sarrion-Perdigones *et al.* (2013).

### ***Nicotiana benthamiana* agroinfiltration**

The transient expression assays were carried out through agroinfiltration of *N. benthamiana* leaves. *N. benthamiana* plants were grown for 5 weeks before agroinfiltration in a growth chamber where the growth conditions were 24°C/20°C light/darkness with a 16h/8h photoperiod. The plasmids were transferred to *A. tumefaciens* strain GV3101 by electroporation. Agroinfiltration was carried out with bacterial cultures grown overnight. The cultures were pelleted and resuspended in agroinfiltration solution (10 mM MES, pH 5.6, 10 mM MgCl<sub>2</sub>, and 200  $\mu$ M acetosyringone). After incubation for 2 h at room temperature with agitation, the optical density of the bacterial cultures was adjusted to 0.1 at 600nm and mixed for co-infiltration with the silencing suppressor P19. Agroinfiltrations were carried out through the abaxial surface of the three youngest, fully expanded leaves of each plant with a 1 ml needle-free syringe.

### **RNA isolation and RT-qPCR gene expression analysis**

Total RNA was isolated from 100 mg of agroinfiltrated leaves using the Gene Jet Plant Purification Mini Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The timing for sample collection was 4 days post infiltration (dpi) for testing individual genes activation and for naringenin optimization and 5 dpi in the case of AP-N3, AP-E, AP-K, AP-Q activation. Before cDNA synthesis, total RNA was treated with the

rDNase-I Invitrogen Kit according to the manufacturer's instructions. An aliquot of 1 µg of DNase-treated RNA was used for cDNA synthesis using the PrimeScript™ RT-PCR Kit (Takara) in a final volume of 20 µl according to the manufacturer's indications. Expression levels for each gene were measured in triplicated reactions, in the presence of a fluorescent dye (SYBR® Premix Ex Taq) using Applied biosystem 7500 Fast Real-Time PCR system with specific primer pairs (Supplementary Table 12). The qPCR conditions employed are those recommended by the manufacturer (TB Green® Premix Ex Taq - Takara): Hold stage 1 cycle 95°C 30 sec, Cycling Stage 40 Cycles 95°C 3 sec and 60°C 30 sec. The qPCR oligo design was performed following the recommendations of the Takara manufacturer. Efficiency curves for each qPCR primer pair were performed using a serial dilution of the template. The calculated efficiency is included in Supplementary Table 12. The *F-box* gene was used as an internal reference (Liu *et al.*, 2012). Basal expression levels were calculated with samples agroinfiltrated with dCasEV2.1 in combination with a non-specific gRNA. mRNA fold change calculations for each sample were carried out according to the comparative  $\Delta\Delta\text{CT}$  method (Livak and Schmittgen, 2001).

### **Liquid chromatography (LC) and ESI mass spectrometry (MS) for flavonoids content analysis**

Leaf samples of three different plants agroinfiltrated with each construct were collected at 4 and 5 dpi and used as triplicates for metabolomics analyses. The same tissue was used for transcriptomics and metabolomics analyses. The tissue was frozen in liquid N<sub>2</sub> and powdered with a grinding mill and, finally, lyophilized. Thirty mg of dried tissue were extracted at 4°C with 1 ml of 30% methanol containing 0.01% formic acid. The preparation was homogenized with a grinding mill and kept on ice for 20 min. After that, the samples were centrifuged at 15000 rpm for 15 min. The supernatant was collected and filtered via a 20-micron cellulose strainer (Regenerated Cellulose Filter, Teknokroma). Three independent biological and two technical replicates per sample were analysed. 20 µl of each sample were injected into an Acquity UPLC system (Waters) coupled to a hybrid quadrupole time-of-flight instrument (QTOF MS Premier). Separation was performed using an HPLC SunFire C18 analytical column with a particle size of 5 µm, 2.1 × 100 mm (Waters). A gradient of methanol and water containing 0.01%

formic acid was used. The gradient started with 95% aqueous solvent and a flow of 0.3 mL per minute. The gradient reached 50% of aqueous solvent at 8 min, increasing the level of organic solvent to 95% at 12 min. The gradient was kept in isocratic conditions for 1 min and later returned to initial conditions in 2 min. The column could equilibrate for 3 min, for a total of 22 min per sample.

Data were collected in MS and MS/MS mode to gain structural information of the detected metabolites. The MS/MS function was programmed in a range of 5 to 45 eV t-wave to obtain each analyte fragmentation spectrum (Pastor *et al.*, 2018). The electrospray ionization was performed in positive and negative mode and analysed individually in order to obtain the best profile of the flavonoid metabolites following the specifications described by Gamir *et al.* (2012).

For unequivocal metabolite determination, samples of naringenin and eriodictyol chemical standard (Sigma) were analysed under the same conditions. The exact mass, specific retention time, and spectrum fragmentation of naringenin and eriodictyol standard were compared to the fragmentation profiles of each sample as described by Schymanski *et al.* (2014).

## **Naringenin content data analysis and statistics**

The raw data obtained were processed by Masslynx 4.1 software and transformed to .cdf files. The positive (ESI+) and negative (ESI-) signals were analysed separately. Naringenin identification was carried out by comparison with a purified standard also analysed in the same conditions.

The quasimolecular ion with  $m/z$  271.06 in ESI negative mode, retention time (5.47 min.), and the fragmentation ions  $m/z$  151 and  $m/z$  119 allowed to unequivocally identify naringenin in the samples (Supplementary Figure 3).

Metabolite amounts were quantified based on the normalized peak area units relative to their respective dry weight. All statistical analyses were conducted with Statgraphics Centurion software for the ANOVA statistical analysis ( $p < 0.05$ ) and means were expressed with the standard error.

## Untargeted data analysis and statistics

The raw data obtained were processed and transformed to .cdf files. The negative (ESI-) signals were analysed employing the xcms software ([www.xcmsonline.scripps.edu](http://www.xcmsonline.scripps.edu)) for comparing all the samples. TIC normalization was applied to each biological triplicate with a baseline of 20 peak intensity relative units. All peaks with a signal lower than 300 in all samples were eliminated to reduce the background. The data analysed comprises the retention time between 1 to 6.5 min (corresponding to the elution conditions for phenolic compounds) and  $m/z$  values ranging from 200 to 1000. Finally, a cut-off of 75% in the coefficient of variation was applied between biological triplicates.

The data obtained were analysed using the MetaboAnalyst5.0 Software ([www.metaboanalyst.ca](http://www.metaboanalyst.ca)). Logarithmic transformation and Pareto scaling were employed as normalization to elaborate the Principal Component Analysis and the Hierarchical Clustering and Heatmap. Euclidean distance and Ward Clustering algorithm were applied as parameters for elaborating Hierarchical Clustering, and an ANOVA test was the statistical method used for generating the 100  $m/z$  significantly different in each group.

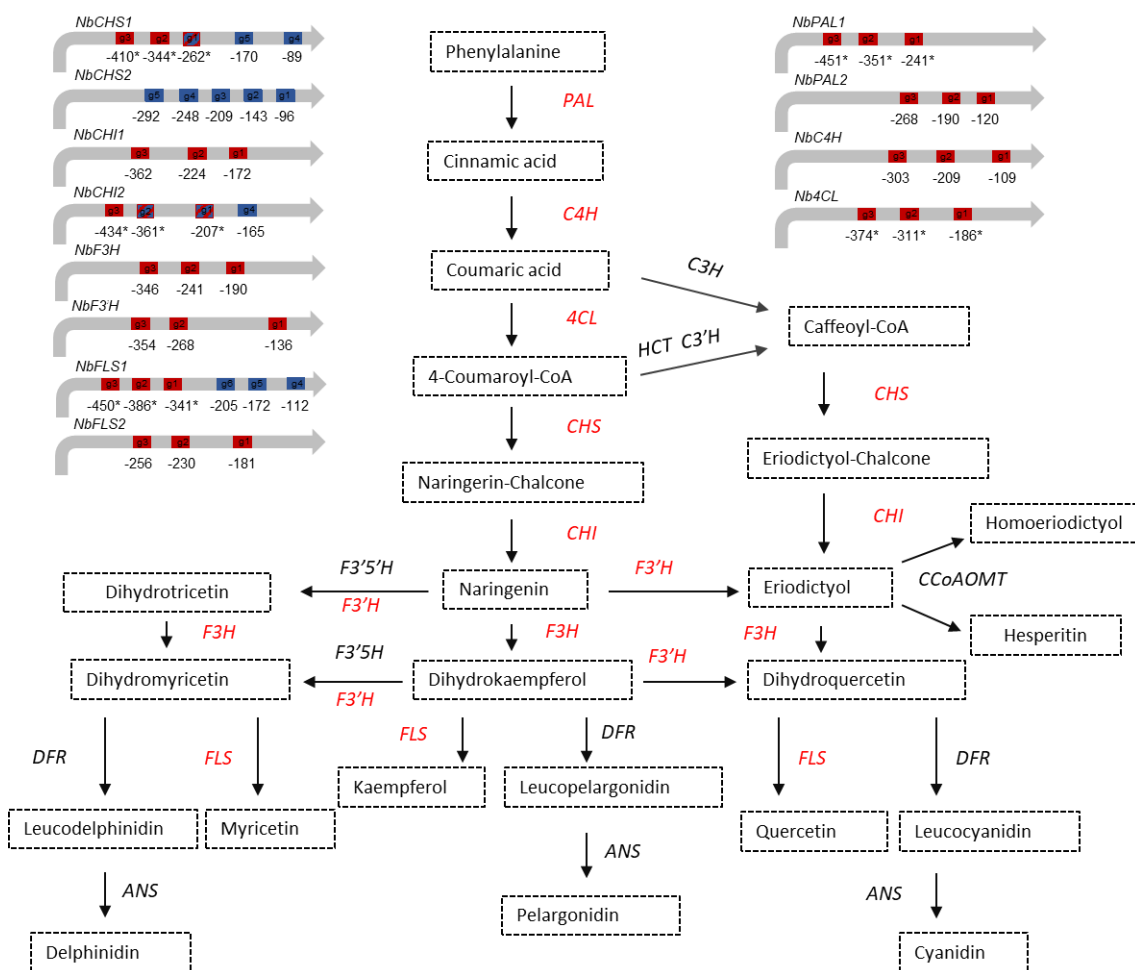
The  $m/z$  values obtained as significantly different in each group were manually clustered into single metabolites employing the original chromatograms. Finally, the tentative identification of each metabolite was carried out using external databases ([www.pubchem.ncbi.nlm.nih.gov](http://www.pubchem.ncbi.nlm.nih.gov)) and the information obtained with their fragmentation profiles and collected in Supplementary Table 13. The quantification of the metabolites was performed employing the parental ion identified in the first fragmentation (F1).



## RESULTS

### Optimization of activation programs for individual genes in the flavonoid pathway

To engineer the flavonoid biosynthesis through the custom upregulation of endogenous genes, the first step consisted in the identification of flavonoid biosynthetic genes in *Nicotiana benthamiana*, including also those encoding upstream enzymes belonging to the general phenylpropanoid pathway. The KEGG reference database with the complete flavonoid pathway was used as a guide for the selection of all candidate genes ([www.genome.jp/kegg/pathway.html](http://www.genome.jp/kegg/pathway.html)). A schematic representation of the pathway can be found in Figure 1, where the enzymatic steps intended for transcriptional activation in this work are highlighted. Candidate gene identification in the *N. benthamiana* genome was carried out manually by homology search using orthologous proteins from *A. thaliana*, *Solanum lycopersicum*, and *Nicotiana tabacum* available in Uniprot ([www.uniprot.org/](http://www.uniprot.org/)) and Solgenomics database ([www.solgenomics.net](http://www.solgenomics.net)). The retrieved candidates were contrasted with the automatic annotation of version v3.3 of the *N. benthamiana* genome assembly ([www.nbenth.com](http://www.nbenth.com)). The allotetraploid nature of *N. benthamiana* results in several paralogs and homoeologues for each enzymatic step in the pathway, therefore the selection of candidates for transcriptional activation was performed according to the following criteria: (i) maximum homology levels with the reference proteins; (ii) completeness of gene annotation, with reliable identification of TSS, a critical parameter that determines the region in which activation efficiency is maximum, usually between nucleotides -100 to -300 from it; (iii) optimal sequence features for gRNA design in the activation region, with an absence of putative off-targets in the *N. benthamiana* genome. On some occasions, discrepancies between *in silico* gene annotation and transcriptomic information were found. In these cases, transcriptomic data was prioritized. In total, twelve different candidate genes were selected for upregulation in two optimization rounds (see below), covering eight enzymatic steps, namely PAL, CHS, CHI, C4H, 4CL, F3'H and F3H, and FLS.



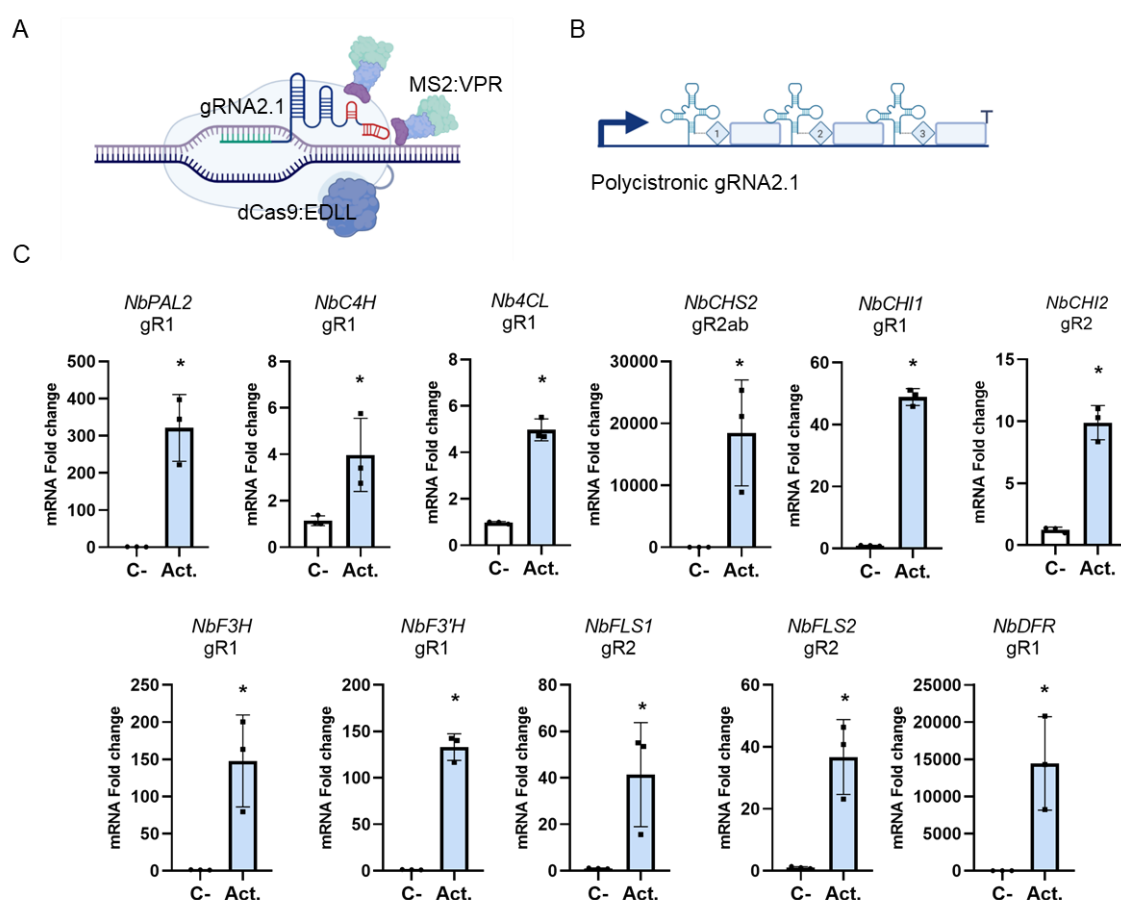
**Figure 1: Schematic representation of the flavonoid biosynthetic pathway in plants.** The different metabolites are represented in boxes. The genes that integrate the pathway are: *ANS* (Anthocyanidin synthase), *CHS* (Chalcone synthase), *CHI* (Chalcone isomerase), *CCoAOMT* (Caffeoyl-CoA O-methyltransferase), *C3H* (4-coumarate 3-hydroxylase), *C3'H* (4-coumaroyl shikimate/quinate 3'-hydroxylase), *C4H* (Cinnamate 4-hydroxylase), *4CL* (4-coumaroyl CoA ligase), *DFR* (Dihydroflavonol 4-reductase), *F3H* (Flavanone 3-hydroxylase), *F3'H* (Flavonoid 3'-hydroxylase), *F3'5'H* (Flavonoid 3' 5-hydroxylase), *F3'-5'H* (Flavonoid 3' 5'-hydroxylase), *FLS* (Flavonol synthase), *HCT* (4-hydroxycinnamoyl CoA: shikimate/quinate hydroxycinnamoyltransferase), *PAL* (Phenylalanine ammonia-lyase). The genes marked in red are those selected for transcriptional activation with dCasEV2.1 in this work. The metabolites targeted for accumulation are shown in bold. A representation of each candidate gene promoter with the targeted gRNA positions is also shown. gRNAs assayed in the first round of optimization are shown in red, and those assayed in the second round are shown in blue. Red/blue striped boxes represent activation sites that were included in activation programs assayed in both optimization rounds. The asterisks represent the gRNA position recalculated with updated information of the TSS available in databases (<https://www.nbent.com/>).

Once the candidate genes were selected, their transcriptional activation was assayed individually in *N. benthamiana* transient assays employing the dCasEV2.1 activation tool (Figure 2A). In each assay, the activation tool was completed with the co-transformation of the target-specific module, which consists of one or two polycistronic gRNAs carrying three target-specific protospacers plus an MS2-binding RNA 2.1 aptamer separated by processable tRNA spacers (Figure 2B). All protospacers were designed against the 5' untranslated region of genes between -100 and -300 bp from the TSS. The list of the gRNAs target sites designed for each gene is listed in Table 1 and depicted in Figure 1. Four days post infiltration (dpi) the samples were collected, and levels of transcriptional activation were evaluated by RT-qPCR. The reference sample used as a negative control was also infiltrated with dCasEV2.1 loaded with a non-target gRNA of *N. benthamiana* genome for the best comparison.

A first round of optimization was carried out for individual gene activation with ten selected genes, namely *NbPAL1*, *NbPAL2*, *NbC4H*, *NbCL4*, *NbCHS1*, *NbCHI2*, *NbF3H*, *NbF3'H*, *NbFLS1* and *NbDFR*. The *NbDFR* gene previously assayed in our group was also added to the analysis for comparison. The multiplexed gRNAs employed in this first attempt are depicted in red in Figure 1. Most endogenous target genes showed remarkable upregulation upon dCasEV2.1 activation treatment (see Supplementary Figure 2A). Surprisingly, the activation program designed for *NbPAL1* resulted in a modest four-fold upregulation, while its homeologue *NbPAL2* showed a 200-fold activation. This was a consequence of erroneous identification of the TSS in *NbPAL1*, which generated the design of the gRNAs in sub-optimal positions. Consequently, *NbPAL2* was selected in this work for further attempts to activate the phenylalanine ammonia-lyase. Unfortunately, the genes *NbCHS1*, *NbCHI2*, and *NbFLS1* showed low activation rates in the first set of experiments (Supplementary figure 2). For this reason, new activation attempts were carried out in a second round of experiments for those enzymatic steps showing suboptimal upregulation, either by introducing new gRNAs targeting the same genes as in round 1, or by targeting alternative genes catalysing the same enzymatic step. The results of the new activation attempts are shown in

---

Supplementary figure 2B. In this second round, improved activation rates were obtained for CHS, CHI and FLS, and their respective gRNA constructs were added to the final set of activation programs to be employed in subsequent experiments. The final set of activation programs for each gene was then re-evaluated in a confirmatory experiment represented in Figure 2C, after the second optimization process, a successful transcriptional activation (> 10-fold) was obtained for all selected enzymatic steps except for *NbC4H* and *Nb4CL*, where only modest four-fold activation rates were obtained. The best activation results were achieved for the previously described *NbDFR* (15000-fold) and *NbCHS2* (18000-fold).



**Figure 2: dCasEV2.1-mediated transcriptional activation of individual genes of the flavonoid pathway in *N. benthamiana*.** A) Schematic representation of the dCasEV2.1 activator comprising the dCas9 fused to the EDLL activation domain in the C-terminus and the coat protein of the phage MS2 fused to the activation domain VPR. B) Schematic representation of the polycistronic gRNA2.1 array including tRNA-spaced gRNAs with MS2-binding RNA aptamers in the 3' end of the scaffold. C) mRNA fold change at 4 dpi obtained

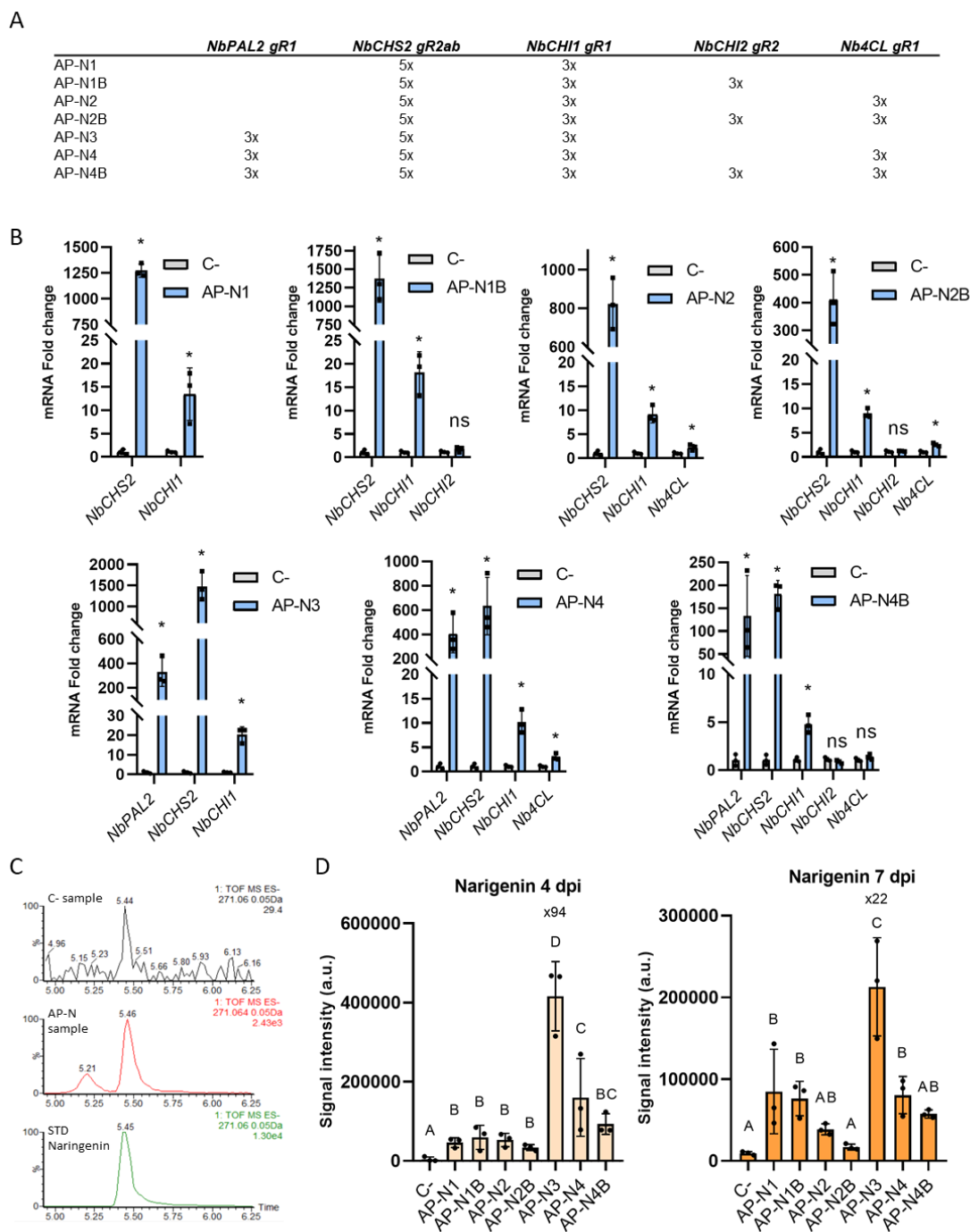
by targeting the endogenous genes with optimized gRNAs in *N. benthamiana* leaves. The gR1 or gR2 indicate the round of optimization where the gRNAs were selected. The *NbCHS2* gR2ab activation was performed with 2 sets of gRNAs. The *NbDFR* gR1 is included as an internal control of activation. Bars represent average fold change  $\pm$  SD (n=3). Asterisks indicate T student significant values (\* =  $p < 0.05$ ). Images were created with BioRender.com.

## Programming naringenin accumulation with dCasEV2.1

The results obtained with the individual activation programs prompted us to undertake the simultaneous activation of several genes in the pathway following a modular polycistronic gRNA strategy (Lowder *et al.*, 2017; Hashimoto *et al.*, 2018). As a first step, we aimed at the co-activation of the enzymatic steps leading to the accumulation of naringenin, therefore involving the genes *NbPAL2*, *NbC4H*, *Nb4CL*, *NbCHS2*, *NbCHI1*, and *NbCHI2*. To select the best dCasEV2.1 activation program possible, we assayed seven different gRNA multiplexing arrangements targeting all the above-mentioned genes except *NbC4H*. Each combination comprised between three to six U6-26-driven polycistronic TUs (Figure 3A). *NbC4H* was omitted as a target gene in all gRNA combinations due to its modest activation rates. All seven combined activation programs were transiently assayed in *N. benthamiana* leaves following the same methodology described for previous experiments. The RT-qPCR results in Figure 3B show that all the assayed programs resulted in significant gene activation, although induction levels were notably reduced compared with programs addressing individual genes. In general, it was observed that smaller multiplexing arrays resulted in higher activation rates of individual genes. For instance, *NbCHS2* activation reached between 800- and 1200-fold with activation programs AP-N1, AP-N1B, AP-N2, and AP-N3 comprising 2 or 3 target genes but dropped below 600-fold in programs targeting four or five genes (AP-N2B and AP-N4B). Following the same trend, those targets showing modest activations in single-gene programs showed even lower inductions with complex activation programs, dropping below significance levels in some cases such as *NbCHI2* when treated with AP-N2B or AP-N4B.

In order to see if the changes in transcript profiles have resulted in the expected metabolic changes, naringenin levels in treated samples were analysed by LC-MS (UPLC-(ESI)-QTOFMS) at 4 and 7 dpi using a purified commercial standard for identification (Figure 3C). As shown in Figure 3D, naringenin content was enriched in all samples as compared with a control leaf where dCasEV2.1 was loaded with an unrelated program. Maximum levels of naringenin were obtained with AP-N3, which targeted *NbPAL2*, *NbCHS2*, and *NbCHI1* simultaneously. In this combination, the levels of the target

compound were raised almost 100-fold as compared with the levels in control samples. The upregulation of *NbPAL* seemed important for the early activation of the pathway since the best levels of naringenin accumulation were found in samples where this enzyme was upregulated. The inclusion of *NbCL4* and/or *NbCHI2* genes in AP-N1B, AP-N2, AP-N2B, AP-N4 and AP-N4B, failed to have a positive effect on naringenin accumulation. On the contrary, the small activation rates achieved in these two individual genes, coupled to an increase in the multiplex gRNA complexity, seem to end up in less efficient activation programs (compare AP-N3 with AP-N4 and AP-N4B), with lower activation levels of *NbPAL2*, *NbCHS2* and *NbCHI1* genes, and consequently, lower levels of naringenin accumulation. The accumulation trends of the targeted metabolite were similar at 4 and 7 dpi, although a drop in signal intensity was observed at the later time point, particularly in samples AP-N3, AP-N4 and AP-N4B.



**Figure 3: Optimization of naringenin production in *N. benthamiana* leaves through dCasEV2.1 activation.** A) Representation of the different activation programs (APs) tested for activating naringenin production (AP-N1, AP-N1B, AP-N2, AP-N2B, AP-N3, AP-N4, and AP-N4B). In the table, Nx values represent the number of guides employed for targeting each gene. B) mRNA fold change at 4 dpi obtained by targeting the endogenous genes of the flavonoid pathway with optimized gRNAs in *N. benthamiana* leaves. Bars represent average fold change  $\pm$  SD (n=3). C) Identification of naringenin in *N. benthamiana* leaves by comparison with a true naringenin commercial standard (STD naringenin) using the ion with



---

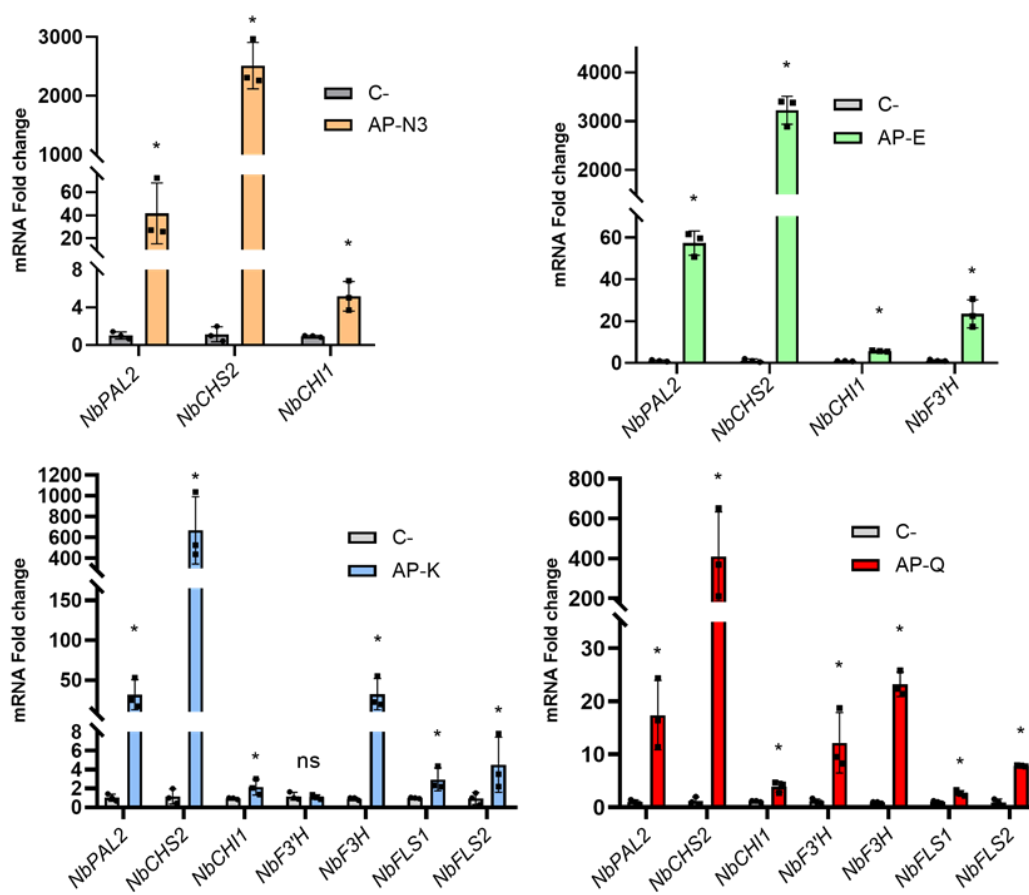
m/z 271.06 in ESI negative mode. C-sample: negative control; AP-N sample: dCasEV2.1-activated sample. D) Relative naringenin quantification in the indicated samples at 4- and 7-dpi. Bars represent average intensity. Different letters indicate statistically significant differences (ANOVA, Tukey HSD test;  $p < 0.01$ ,  $n = 3$ ). Asterisks indicate T student significant values (\* =  $p < 0.05$ ).

## Customization of flavonoid composition as a study case for single metabolite precision level engineering

Despite the indications that complex gRNA programs resulted in lower activation efficiencies than simpler ones, the remarkable levels of naringenin accumulation obtained with shorter programs suggested that there was still room for adding new instructions to AP-N3, thus extending regulation further downstream in the pathway. Furthermore, naringenin constitutes a crossroad point from which several compounds can be derived depending on the set of downstream enzymes to be activated. Thus, the activation of *F3'H* on top of AP-N3 would convert naringenin into a different flavanone, eriodictyol. Moreover, naringenin can be used as starting point for the accumulation of two important flavonols, kaempferol and quercetin. Steering the metabolic flux to produce kaempferol would require the simultaneous activation of two enzymes, first *F3H* to produce dihydrokaempferol, and next *FLS* to introduce a double bond in the C ring, yielding kaempferol. Alternatively, the production of quercetin can be induced by taking the eriodictyol program as a basis, and adding activation instructions for *F3H* and *FLS*, producing the first dihydroquercetin and finally quercetin. Following this rationale, three new metabolite-specific gRNA programs (Figure 4A), namely AP-K, AP-E, and AP-Q, were constructed and assayed in *N. benthamiana* next to AP-N3 to produce kaempferol, eriodictyol, quercetin, and naringenin, respectively. Figure 4B shows the

A

	<i>NbPAL2</i> gR1	<i>NbCHS2</i> gR2ab	<i>NbCHI1</i> gR1	<i>NbF3'H</i> gR1	<i>NbF3H</i> gR1	<i>NbFLS1</i> gR2	<i>NbFLS2</i> gR2
AP-N3	3x	5x	3x				
AP-E	3x	5x	3x	3x			
AP-K	3x	5x	3x		3x	3x	3x
AP-Q	3x	5x	3x	3x	3x	3x	3x



**Figure 4: Design and relative expression analyses of activation programs (APs) for naringenin, eriodictyol, kaempferol, and quercetin production in *N. benthamiana*.** A) Representation of the Activation Programs (APs), the gRNAs and the targeted genes included in each of them (AP-N3 for Naringenin Activation Program, AP-E for Eriodictyol Activation Program, AP-K for Kaempferol Activation Program, and AP-Q for Quercetin Activation Program). In the table, Nx values represent the number of guides employed for targeting each gene. B) mRNA fold change at 5 dpi obtained by targeting the endogenous genes with optimized gRNAs in *N. benthamiana* leaves for the groups AP-N3, AP-E, AP-K, and AP-Q. Bars represent average fold change  $\pm$  SD (n=3). Asterisks indicate T student significant values (\* =  $p < 0.05$ ).

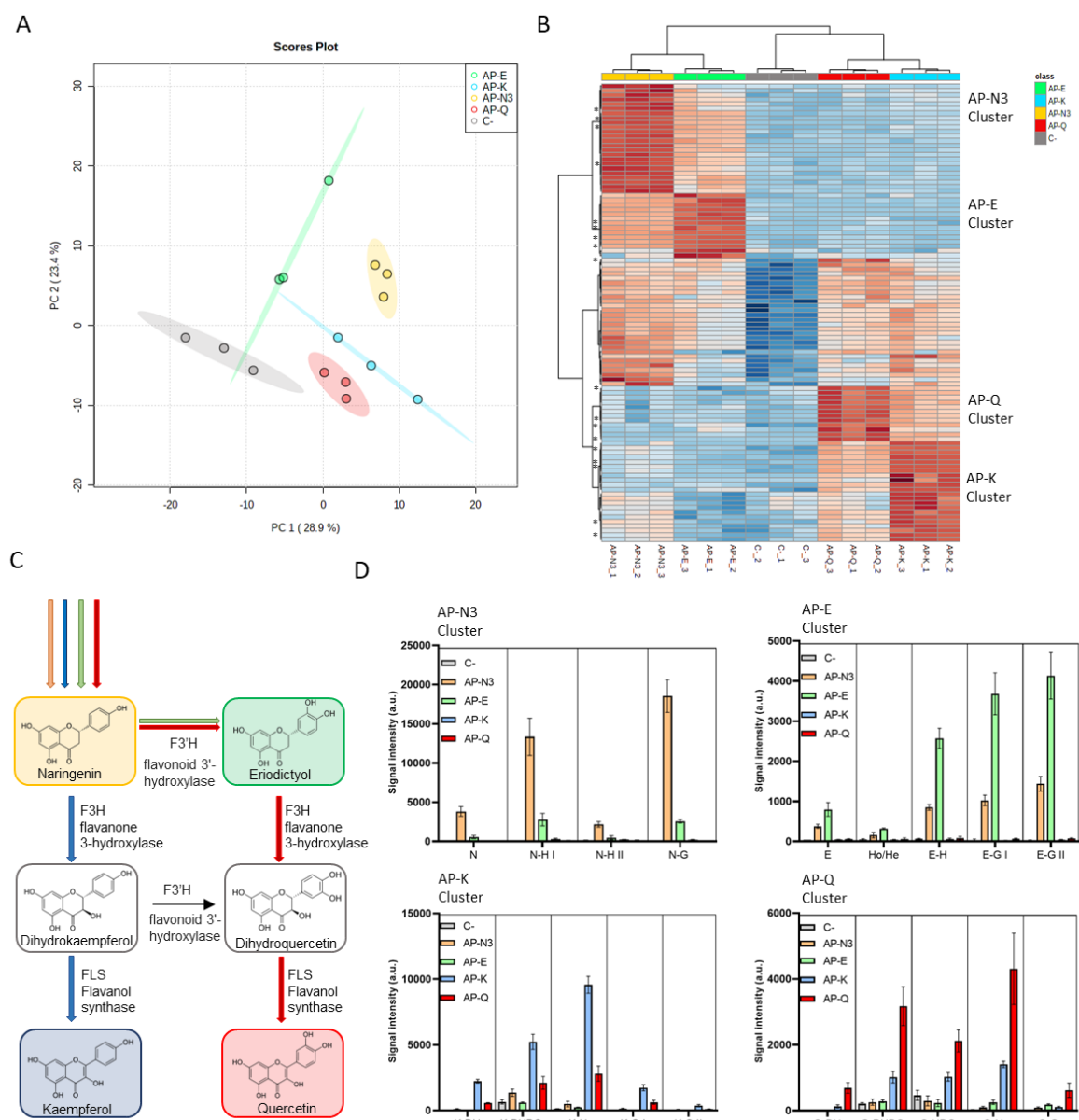
gene activation profiles obtained for each combination, where it can be observed that an increase in program complexity has a negative effect on the overall induction levels compared to simpler programs. This was clearly observed in the activation levels of *NbCHS2*, which drops by approximately 5-fold when more than four genes are targeted simultaneously. Despite the progressive reduction in activation levels, a significant upregulation is still observed in all enzymes assayed even in the more complex program (AP-K and AP-Q). It is worth noticing that *NbF3'H* is only activated in AP-Q but not in AP-K, with this serving as an additional indication of the specificity of the activation programs and discarding positive feedback as a cause for the observed upregulations.

To understand to what extent the customized activation was specific for the intended metabolic steps, leading to differential flavonoid composition in treated leaves, an untargeted LC-MS metabolomic analysis was carried out with the same samples previously analysed by RT-qPCR. As the Principal Component Analysis (PCA) in Figure 5A indicates, each activation program produced a distinct and characteristic metabolite profile, with a first main component separating controls from flavonoid-activated samples accounting for 28.9% of the variance, and a second principal component separating flavanones Activation Programs to the flavonols Activation Programs accounting for 23.4% of the variance. The level of precision achieved with the four activation programs became even more evident when the 100 most significantly different features were hierarchically clustered (Figure 5B). Here, a perfect clustering was observed that parallels the activation programs and the control sample. A detailed version of the heatmap can be found in Supplementary Figure 4. The differential *m/z* ions and their respective retention times are listed in Supplementary Table 13. Furthermore, when metabolites in each cluster were tentatively identified attending to their retention time and the characteristic *m/z* ratios (see Materials and Methods), a remarkable match was found between the activation program employed and the predominant metabolites in the samples. Thus, as expected, AP-N3 samples accumulate naringenin aglycon and three other glycoside derivatives more than any other sample. Eriodictyol and its sugar conjugates are the metabolites accounting for the AP-E cluster, although certain levels of eriodictyol are also found in the AP-N3 cluster. This is not

surprising

since

both



**Figure 5: Analysis of the metabolic profiles of *N. benthamiana* leaves treated with activation programs (APs) for naringenin, eriodictyol, kaempferol, and quercetin.** A) Principal Component Analysis resulting from the untargeted LC-MS data obtained in ESI negative mode from leaves treated with programs AP-N3, AP-E, AP-K, and AP-Q and control samples (C-) agroinfiltrated with unprogrammed dCasEV2.1. B) Hierarchical cluster analysis and heatmap representation of the Control, AP-N3, AP-E, AP-K, and AP-Q metabolic profiles, with three biological replicates per condition (ESI-). The m/z represented are the 100 most significant using an ANOVA analysis ( $p < 0.05$ ). The data was obtained using Euclidean

distance and Ward's minimum variance method. Red indicates up-regulated and blue downregulated features. The asterisk symbols represent the m/z selected for quantifying the metabolites in Figure 5D. C) Schematic representation of selected target metabolites of the flavonoid pathway and the genes involved in each enzymatic reaction required for their biosynthesis. The coloured bars represent the targeted gene for each group (AP-N3 in orange, AP-E in green, AP-K in blue, and AP-Q in red). D) Relative abundance of identified metabolites. The ions employed for the metabolic quantification are the parental ions identified in the heatmap as single compounds (See also Supplementary Table 13). N, Naringenin; N-H, Naringenin Hexose; N-G, Naringenin Glucoside; E, Eriodictyol; E-H, Eriodictyol Hexose; E-G, Eriodictyol Glucoside; K-DH, Kaempferol Dihexose; K-DHDO, Kaempferol Dihexose-deoxyhexose; K-H, Kaempferol Hexose; K-G, Kaempferol Glucoside; Q-DH, Quercetin Dihexose; Q-DHDO, Dihexose-deoxyhexose; Q-HDO, Quercetin Hexose-deoxyhexose; Q-H, Quercetin Hexose; Q-G, Quercetin Glucoside.

flavanones are only one enzymatic step away from each other, and eriodictyol can be also synthesized from coumaric acid following a secondary branch in the pathway (see Figure 1). Remarkably, flavonols are almost completely absent both in AP-N3 and AP-E. On the contrary, kaempferol and its glycosylated derivatives are the main differentially accumulated compounds in AP-K samples, and conversely, the quercetin aglycon and its conjugated derivatives are most abundant in AP-Q samples. As could be anticipated, a certain level of cross-contamination is observed in both flavonol programs. Again, this is not entirely unexpected as both AP-K and AP-Q share 5 out of 6 steps in their respective programs. It is worth noting that flavonol levels remain low in AP-K and AP-Q samples, indicating that successful activation of downstream genes is responsible for the specific accumulation of flavonol compounds.

## DISCUSSION

Programmable Transcriptional Activators (PTAs) based on CRISPR/dCas9 architecture are powerful tools for the transcriptional regulation of a wide spectrum of targets. After the first generation of PTAs based on the direct translational fusion of TADs to dCas9 showed modest induction activities (Piatek *et al.*, 2015), subsequent generations have emerged incorporating multiple AD anchoring sites such as multi-epitope chains (Papikian *et al.*, 2019), RNA aptamers, or combinations of them. These upgraded

versions considerably boosted PTAs ability to produce strong activation of targeted genes (Lowder *et al.*, 2018a; Lee *et al.*, 2019). These improvements, added to the multiplexing capacity of CRISPR/dCas9, have turned CRISPR-PTAs into extraordinary tools for Plant Synthetic Biology applications. Of particular interest is the ability offered by CRISPR-PTAs to tinker with enzyme expression in metabolic pathways, thus shaping the plant metabolite composition. In this work, we tinkered with a section of the flavonoid pathway, exploring the limits of the tool, to find out that dCasEV2.1, the CRISPR-PTA developed in our laboratory, offers an unprecedented level of precision in plant metabolic engineering interventions. We anticipate that other CRISPR-PTAs developed by other groups, having an equivalent mode of action and showing similar gene activation levels, would perform similarly well when used for the same purposes (Pan *et al.*, 2021b). We show that with a careful selection of gRNAs, simultaneous regulation of up to six enzymatic steps can be achieved, leading to the highly preferential accumulation of individual flavanones (naringenin or eriodictyol) and flavonols (quercetin or kaempferol). The ability to harness a pathway towards the production of a single predominant metabolite has important implications in plant biomanufacturing and biorefinery since the isolation of pure valuable metabolites is known to be hindered by the presence of contaminant compounds belonging to the same pathway, hence showing similar chemical properties. In the examples described here, a considerable proportion of the targeted metabolites were found in glycosylated forms. *N. benthamiana*, like many other plant species, is promiscuous in glycosyltransferase activities (Wang, 2009). From a biomanufacturing point of view, glycosylated forms constitute a relatively minor problem due to the affordability of industrial glycosylases. As an alternative, glycosyltransferases could be also targeted for programmable repression, leading to the predominant accumulation of the aglycon.

In this work, we only use transcriptional activators as tools to re-route metabolic fluxes. Programmable repression is partially dispensable for the control of the flavonoid pathway because in *N. benthamiana* leaves the flux through the pathway is low. However, the interplay of programmable repressors would enable further refining of the accuracy of re-routing programs. As mentioned above, the repression of glycosyltransferases would serve to ensure the predominance of aglycon forms if

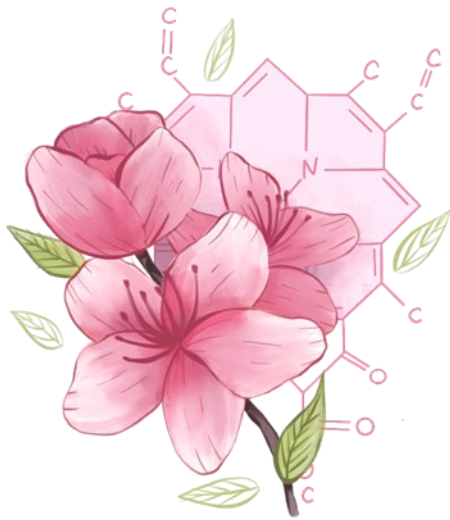
required. Furthermore, as shown in Figure 1, eriodictyol can be synthesized by an alternative route that involves the action of *NbC3'H*, *NbC3H*, and *NbHCT* genes. Programmed repression of one or more of these genes could contribute to reducing the levels of accompanying metabolites when the production of a single flavonol is attempted. Furthermore, programmable repression will be strongly required for harnessing metabolic pathways other than the flavonoid pathway which are highly active by default. Unfortunately, there are few examples in the literature showing highly efficient repression based on CRISPR architecture in plants (Tang *et al.*, 2017). All currently described tools, although helpful, are probably insufficient in providing full control due to their inability to conduct strong transcriptional repressions in highly active genes. It has been proposed that dCas9, due to its mode of action when binding DNA, which implies a relaxation of the chromatin in the surrounding area, has limited capacity to act as a strong repressor when fused to conventional repressor domains. This limitation might be circumvented by the use of epigenetic repressors adding reversible chromatin silencing marks (Gallego-Bartolomé, 2020). Alternatively, programmable repression could be achieved by other means, such as post-transcriptional gene silencing tools (Mahas *et al.*, 2019). Given the remarkable ability of dCasEV2.1 to program transcriptional activation, its combination with efficient repressors would enable near full control of metabolic pathways in plants.

For the delivery of activation programs, we made use of transient *Agrobacterium*-based transformation. Transient delivery of genetic information in the form of T-DNA or RNA is becoming increasingly popular in plant biotechnology. Recently, transient reprogramming of crop plants was shown using RNA virus-based delivery systems, either as viral particles or mediated by *Agrobacterium* (Torti *et al.*, 2021). Agroinfiltration has become not only a widely used experimental procedure (Norkunas *et al.*, 2018; Grosse-Holz *et al.*, 2018) but also a potent and scalable plant biomanufacturing strategy as recently demonstrated with the production of plant-made vaccines against influenza and SARS-CoV2 (D'Aoust *et al.*, 2008; Diego-Martin *et al.*, 2020). On the other hand, the gRNA elements in PTAs can be also transiently delivered using viral vectors, as recently shown using TRV as a delivery agent (Ghoshal *et al.*, 2020). Reprogramming metabolic pathways using transient tools has the additional advantage of circumventing the need

for stable transgenics, giving regulatory advantages in some areas. Alternatively, the stable transformation of the activation programs could also be envisioned as a powerful tool in crop breeding. Stable integration in the plant's genome could circumvent the complexity limits evidenced in this work, which seem to occur when a large number of gRNAs are encoded in a single construct. This effect could be attributed to the cargo capacity of the T-DNAs, or the plant's ability to cope with highly repetitive gRNA structures. Also, at some point, an increasing number of gRNAs targets must turn dCas9 into the limiting factor. Stable transgenic programs could be pyramided in different genomic loci, avoiding limits imposed by T-DNA cargo capacity and repetitive structures. If required, the expression of the constant components in dCasEV2.1 could be boosted with state-of-the-art strategies, preventing them from becoming a limiting factor for activation (Pasin *et al.*, 2017). Connecting integrated programs with endogenous or exogenous cues using appropriate inducible/conditional systems (Ochoa-Fernandez *et al.*, 2020; Randall, 2021) would provide the ultimate ability to customize plant metabolic composition using its endogenous metabolic pathways, while avoiding detrimental effects or due to continuous activation.

In summary, we show here that CRISPR/dCas9-based transcriptional activators provide sufficient precision and multiplexing capacity to reprogram metabolic pathways and customize metabolic composition in plants. This ability has important implications in feed and food design, as well as in the valorisation of industrial crops.





## Chapter 3

**Potato virus X -delivered CRISPR activation programs lead to strong endogenous gene induction and transient metabolic reprogramming in *Nicotiana benthamiana*.**

**S Selma**, S Gianoglio, M Uranga, M Vázquez-Vilar, A Espinosa-Ruiz, M Drapal, PD Fraser, JA Daròs, D Orzaez (2022), Potato virus X -delivered CRISPR activation programs lead to strong endogenous gene induction and transient metabolic reprogramming in *Nicotiana benthamiana*

bioRxiv 2022.04.21.489058; DOI: <https://doi.org/10.1101/2022.04.21.489058>

My contribution to this work was essential to this publication. I contributed to most of the analysis performed and a major part of the manuscript writing.



## ABSTRACT

Programmable transcriptional regulators based on CRISPR architecture are promising tools for the control of plant gene expression. In plants, CRISPR gene activation (CRISPRa) has been shown effective in modulating development processes, such as the flowering time, or customising biochemical composition. The most widely used method for delivering the CRISPR components into the plant is *Agrobacterium tumefaciens*-mediated genetic transformation, either transient or stable. However, due to their versatility and their ability to move, virus-derived systems have emerged as an interesting alternative for supplying the CRISPR components to the plant, in particular the gRNA, which represents the variable component in CRISPR strategies. In this work we describe a *Potato virus X* (PVX)-derived vector that, upon agroinfection in *N. benthamiana*, serves as a vehicle for gRNAs delivery, producing a highly specific Virus-Induced Gene Activation (VIGA). The system works in combination with a *Nicotiana benthamiana* transgenic line carrying the remaining complementary CRISPRa components, specifically the dCasEV2.1 cassette, which has previously been shown to mediate strong programmable transcriptional activation in plants. Using an easily scalable, non-invasive spraying method, we show here that gRNAs-mediated activation programs move locally and systemically generating a strong activation response in different target genes. Furthermore, by activating three different endogenous MYB transcription factors, we demonstrate that this PVX-based virus-induced gene reprogramming (VIGR) strategy results in program-specific metabolic fingerprints in *N. benthamiana* leaves characterized by distinctive phenylpropanoid-enriched metabolite profiles.

**Keywords:** CRISPRa, programmable transcriptional regulators, PVX, VIGR, virus induced-gene silencing.

## INTRODUCTION

The emerging CRISPR-Cas systems provide a battery of specific and efficient tools for gene editing and regulation (Mao et al. 2019; Moradpour et al. 2020). The recent achievements in targeted gene regulation in plants employing CRISPR/Cas-based programmable transcription factors not only enhance our ability to accurately explore the links between gene expression and phenotype (Lowder *et al.*, 2018b; Papikian *et al.*, 2019; Pan *et al.*, 2021c) but also open a new paradigm for targeted crop improvement by, for example, tuning flowering time or customizing metabolic composition (Lessard *et al.*, 2002; Charfeddine *et al.*, 2019; Maeda and Nakamichi, 2022). CRISPR-based transcriptional regulators typically comprise a catalytically inactive Cas enzyme with transcriptional regulatory domains anchored to its structure. This allows the specific binding of the chimeric transcriptional regulator to the genomic DNA, guided by the 20 nucleotides of the guide RNAs (gRNAs) (Lee *et al.*, 2019). The most widely used Cas nuclease is the Cas9 endonuclease from *Streptococcus pyogenes*, however other nucleases have been also optimized for achieving transcriptional regulation in plants (Zhang *et al.*, 2021c). Currently, engineering efforts are focused on improving the efficiency and specificity of programmed gene regulation and also on the development of optimized delivery methods for the information encoded in the gRNA sequence. Thus, in an ideal scenario, crop plants could contain integrated CRISPR systems capable of executing gene activation instructions that would be delivered in the form of gRNAs commands that move systemically throughout the plant (Molina-Hidalgo *et al.*, 2020).

Initial approaches for the delivery of CRISPR components to the plant were based on the widely used *Agrobacterium*-mediated T-DNA transformation, biolistic-delivery or protoplast transformation methods (Mathur and Koncz, 1998; Wu *et al.*, 2020). The *Agrobacterium*-mediated transformation relies on already established plant transformation protocols (Tzfira and Citovsky, 2006; Danilo *et al.*, 2019). It is easy to handle and does not present strict limitations on the size of DNA constructs, but it requires the generation of stable transgenics through tissue culture and regeneration. *Agrobacterium*-mediated transformation can operate transiently in some species, such as *N. benthamiana* (Wydro *et al.*, 2006), without generating stable transgenic lines. In

this case, however, the action of the CRISPR elements will be limited to the area of agroinfiltration. The biolistic delivery and protoplast transfection methods provide alternatives for plant species unable to be transformed with *Agrobacterium*, and offer non-transgenic options for genome editing (Hamada *et al.*, 2017), like the preassembling of the Cas protein and the gRNA to form ribonucleoproteins (RNPs) (Liang *et al.*, 2019; Zhang *et al.*, 2021b). However, these delivery strategies present some disadvantages, such as the instability and the rapid degradation of the RNP complex, and do not circumvent the requirements of plant regeneration, which is not suitable for all plant species.

Classic work showed that viruses can be used to silence plant endogenous genes by simply harbouring a sequence fragment homologous to the target gene in an approaches known as virus induced-gene silencing (VIGS) (Lu *et al.*, 2003). In addition, recent studies highlight the potential use of viral vectors as transient delivery vehicles for CRISPR-Cas components in many biological systems including plants (Platt *et al.*, 2014; Senís *et al.*, 2014; Xu *et al.*, 2019). This approach, commonly termed virus-induced genome editing (VIGE), has mainly focused on the delivery of one or more gRNAs using RNA or DNA virus vectors in transgenic plants that stably express the Cas nuclease (Gentzel *et al.*, 2022). The range of viral systems used for gRNA-delivery in VIGE approaches is expanding, also enlarging the range of suitable plant hosts. Recently described VIGE vectors include the *Tobacco rattle virus* (TRV) (Ali *et al.*, 2015; Ellison *et al.*, 2020), the *Cabbage Leaf Curl virus* (CaLCuV) (Yin *et al.*, 2015), the *Tobacco mosaic virus* (TMV) (Cody *et al.*, 2017), the *Pea early browning virus* (Ali *et al.*, 2018), the *Foxtail mosaic virus* (FoMV) (Mei *et al.*, 2019), the *Barley stripe mosaic virus* (BSMV) (Hu *et al.*, 2019; Li *et al.*, 2021) and the *Beet necrotic yellow vein virus* (BNYVV) (Jiang *et al.*, 2020). One of the previously reported VIGE systems that showed remarkable efficiency in terms of gene editing was based on the *Potato virus X* (PVX) (Ariga *et al.*, 2020), a member of the genus Potexvirus that infects 62 plant species, several of them belonging to the Solanaceae family (Lico *et al.*, 2015). In addition, the PVX infection can be easily traced due to the phenotypic alterations produced in the host in a relatively short period. This recombinant virus was engineered to drive the expression of the gRNA by the sub-genomic coat protein (CP) promoter. Also, this viral system entails a multiplexing

strategy for gRNA expression with the particularity that the PVX-based gRNA delivery does not need the presence of a tRNA-processing system for properly expressing several gRNAs in tandem (Uranga *et al.*, 2021a).

Viral vectors have been also proposed as shuttles for the transient delivery of exogenous, RNA-encoded information to plant crops, leading to the ectopic and generalized expression of, for example, added-value recombinant proteins, defence-related genes or developmental regulators (Marillonnet *et al.*, 2005; Gleba *et al.*, 2013; Torti *et al.*, 2021). This inspiring new breeding concept of transient genetic reprogramming can be expanded even further with the introduction of CRISPR programmable regulators. In this newly proposed scheme, gRNAs delivered to the plant using viral vectors would serve to reprogram endogenous gene expression. In the present study, we aimed to explore the potential of PVX-based gRNA delivery for strong customized transcriptional activation in plants, thus expanding the toolbox of viral systems employed for gene regulation, which is currently based only on TRV (Ghoshal *et al.*, 2020; Khakhar *et al.*, 2021). To achieve this, the PVX-based VIGE vector developed by Uranga *et al.* (2021) was further engineered to express tandem repeats of modified gRNAs incorporating two RNA aptamers that bind the coat protein of the MS2 phage. This modification is an adaptation to the so-called dCasEV2.1 system, a strong CRISPR programmable activator earlier described by our group (Selma *et al.*, 2019). In addition, the constant elements of the dCasEV2.1 system were stably transformed in *N. benthamiana*. The constant dCasEV2.1 module comprises two transcriptional units (TUs) driven by the *Cauliflower mosaic virus* (CaMV) 35S promoter: (i) a TU encoding an inactive version of the Cas9 (dCas9) fused to the plant activation domain EDLL (Tiwari *et al.*, 2012), and (ii) a second TU encoding the coat protein of the MS2 phage fused to the VPR activator. VPR is a tandem fusion domain comprising the viral transactivation domains VP64, P65 and Rta (Chavez *et al.*, 2015). Through a spray-based agrodelivery method that is easy to handle, non-invasive, and adapted for large scale applications, we demonstrated that the modified gRNAs were efficiently delivered to the plant, leading to a potent PVX-based virus-induced gene reprogramming (VIGR). By targeting three different endogenous MYB factors (*NbODO1*, *NbMYB24*, and *NbMYB21*), allegedly involved in regulating the phenylpropanoid pathway, we generated three different

phenylpropanoid-enriched chemotypes in *N. benthamiana* leaves, obtaining information on the specific transcriptomic and metabolic fingerprint produced by each MYB factor.

## MATERIALS AND METHODS

### Generation and selection of *N. benthamiana* dCasEV lines

The *N. benthamiana* YFP-dCasEV lines were generated following the transformation protocol described previously (Clemente, 2006). The construct (GB2618) was transferred to the LBA4404 *A. tumefaciens* strain used for plant transformation. Murashige and Skoog (MS) plates supplied with kanamycin at 100 mg/L were used to select the transgenic T0 lines. In addition, a YFP analysis was carried out to double-check the construct incorporation. Transgenic lines were sorted based on the T1 plants capable of generating efficient transcriptional activation. pSIDFR:Luciferase was employed as a reporter for testing the activation rates obtained by incorporating the corresponding gRNA to target the SIDFR promoter (gRNA:SIDFR). A wild type *N. benthamiana* was infiltrated in parallel with the same reporter, the gRNA:SIDFR and dCasEV2.1 constant module construct for a fair comparison of the activation results obtained. Also, the T1 plants were evaluated by segregation analysis for selecting a single-copy T-DNA insertion.

The candidate single copy T2 of L3 and L6 lines were analysed for homozygosity and finally, YFP-dCasEV-L6.4.1 was selected as a homozygous population and kept for the following experiments.

### gRNAs design and viral vector construction

Initial *N. benthamiana* endogenous gene *NbDFR* was selected for CRISPRa optimization. The protospacer was previously described in Selma *et al.* (2019). The design of gRNAs for *NbODO1*, *NbMYB21* and *NbMYB24* were performed employing the Benchling CRISPR application tool ([www.benchling.com](http://www.benchling.com)). The protospacer sequences were selected within the activation window located between -100 and -300 bp upstream of the

Transcriptional Start Site (TSS) for each gene. Protospacer sequences were listed in Supplementary Table 14. Plasmid pPVX was previously described in Uranga *et al.* (2021a). The full-length PVX cDNA (GenBank accession number MT799816) is flanked by the Cauliflower mosaic virus 35S promoter and *A. tumefaciens* Nopaline synthase (Nos) terminator. The PVX CP promoter drives the expression of the heterologous gene and a PVX CP, with a deletion of 29 initial codons (Dickmeis *et al.*, 2014), is driven by a heterologous promoter derived from the Bamboo mosaic virus. The double sgRNA2.1 for CRISPRa were included in recombinant virus plasmids through PCR amplification with high-fidelity Phusion DNA polymerase (Thermo Scientific, Waltham, MA, USA) and Gibson DNA assembly (NEBuilder HiFi DNA assembly master mix, New England Biolabs, Ipswich, MA, USA). Primers employed for sgRNA2.1 construction and adaptation for PVX recombinant plasmids were listed in Supplementary Table 15. All the plasmids were confirmed by Sanger sequencing. As negative control of infection, the PVX::gXT2 (non-specific for activation) was employed without adding the 2.1 aptamers.

### **Plant inoculation**

Transgenic *N. benthamiana* dCasEV2.1 plants were grown in growth chambers at 25°C under a 16/8-h day-night cycle. The inoculation was performed when the plant age was 4 weeks post sowing. The PVX recombinant plasmids were transferred to *Agrobacterium tumefaciens* strain GV3101 by electroporation. Inoculation was carried out with overnight grown bacterial cultures. The cultures were pelleted and resuspended on agroinfiltration solution (10 mM MES, pH 5.6, 10 mM MgCl<sub>2</sub>, and 200 µM acetosyringone) and incubated for 2 h at room temperature with agitation. The optical density of the bacterial cultures was adjusted to 0.5 at 600nm for the syringe inoculation method and 0.02, 0.004 and 0.0008 for the spray inoculation method and 0.04% of surfactant was added. The final optical density employed for the *NbODO1*, *NbMYB21* and *NbMYB24* spray induction was 0.0008. Two leaves per plant were employed in both inoculation methods, using a 1 ml syringe on the abaxial side of the leaf or a complete spray of both sides of the leaves. Control plants were inoculated with PVX::gXT2 following the same procedure. The samples were collected at 9- and 14-days post infiltration (dpi) from the first and second symptomatic leaf employing the syringe



approach. Samples from the youngest sprayed leaf were collected at 5 and 9 dpi to measure the kinetics induction of *NbDFR* in local infection. Samples from the first symptomatic leaf were collected at 9, 14 and 20 dpi to measure the kinetics of the *NbDFR* gene in systemic infection. Finally, samples from the *NbODO1*, *NbMYB21* and *NbMYB24* induction assay were collected 1 at 4 dpi from the first symptomatic leaf. All the samples were immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

### **Luciferase/Renilla activity determination**

The determination of the luciferase/renilla (Fluc/Rluc) activity was carried out by collecting one disc per leaf ( $d = 0.8$  cm, approximately 18–19 mg) at 5 dpi. The samples were homogenized and extracted with 375  $\mu\text{l}$  of 'Passive Lysis Buffer,' followed by 10 min of centrifugation ( $14,000\times g$ ) at  $4^{\circ}\text{C}$ . Then, the supernatant was collected as working plant extract. Fluc and Rluc activities were determined following the Dual-Glo<sup>®</sup> Luciferase Assay System (Promega) manufacturer's protocol with minor modifications: 10  $\mu\text{l}$  of working plant extract, 40  $\mu\text{l}$  of LARII and 40  $\mu\text{l}$  of Stop&Glow Reagent were used. Measurements were made using a GloMax 96 Microplate Luminometer (Promega) with a 2 s delay and a 10 s measurement.

Fluc/Rluc ratios (RPU) were determined as the mean value of three biological replicates coming from three independent agroinfiltrated leaves of the same plant. The RPU were normalized to the Fluc/Rluc ratio obtained for a reference sample, that measures relative transcriptional activities (RTA) of the evaluated promoter fused to the reporter.

### **RNA isolation and qRT-PCR gene expression analysis**

Total RNA was isolated from 100 mg of leaf frozen tissue employing an isolation kit (Gene Jet Plant Purification Mini Kit-ThermoScientific) according to the manufacturer's instructions. RNA was treated with DNase-I Invitrogen Kit and its concentration was adjusted for cDNA reaction. 800ng of total RNA was used for cDNA synthesis using PrimeScript<sup>™</sup> RT-PCR Kit (Takara) in 20  $\mu\text{L}$  final volume according to the manufacturer. Expression levels were measured employing four biological replicates for *NbDFR* induction assays and five biological replicates for *NbODO1*, *NbMYB21* and *NbMYB24*

induction assays. Each sample was measured in three technical replicates reactions, in presence of fluorescent dye (SYBR® Premix Ex Taq), using the Applied biosystem QuantStudio 3 equipment. The specific primers for detection of the target genes were listed in Supplementary Table 16. F-BOX protein was used as an internal reference gene (Liu *et al.*, 2012). Basal expression levels were calculated from the control samples inoculated with PVX::gXT2 recombinant plasmid. Calculations of each sample were carried out according to the comparative  $\Delta\Delta$ CT method (Livak and Schmittgen, 2001).

### **Volatile organic compound analysis and statistics**

70 mg of frozen and ground symptomatic leaf samples were incorporated in a 10 mL headspace screw-cap vial. The samples were prepared by adding 1 mL of 5 M CaCl<sub>2</sub>, 150  $\mu$ l of 500 mM EDTA (pH = 7.5) and a 3  $\mu$ l lavadulol at 10 ppm as an internal reference and sonicated for 5 minutes. Volatile compounds were captured by solid-phase microextraction (HS-SPME) with a 65  $\mu$ m polydimethylsiloxane/divinylbenzene (PDMS/DVB) SPME fibre (Supelco, Bellefonte, PA, USA). Volatile extraction was performed automatically by a CombiPAL autosampler (CTC Analytics). Vials were first incubated at 80°C for 3 minutes with 500 rpm agitation. Then, the fibre was exposed to the vial through the headspace for 20 min, under the same conditions of temperature and agitation. Desorption was performed at 250°C for 1 min (splitless mode) in the injection port of a 6890 N gas chromatograph (Agilent Technologies). After desorption, the fibre was cleaned in an SPME fibre conditioning station (CTC Analytics) at 250°C for 5 min under a helium flow. Chromatography was performed on a DB5ms (60 m, 0.25 mm, 1  $\mu$ m) capillary column (J&W) with helium as the carrier gas at a constant flow of 1.2 ml min<sup>-1</sup>. The oven conditions were 40°C for 2 min, and a 5°C min<sup>-1</sup> ramp was programmed until reaching 300°C with a final hold at 300°C for 5 min. Two blank controls with 1 ml of 5 M CaCl<sub>2</sub> and 150  $\mu$ l of 500 mM EDTA were included in the experiment.

Identification of the volatile compounds was performed by employing a customized library of *N. benthamiana* based on the NIST database (See supplementary data). A non-targeted analysis for the differential emission of volatile compounds in the *NbODO1*, *NbMYB21*, and *NbMYB24* samples was performed. Peak intensities were calculated by employing the Agilent Mass Hunter Workstation Software for Quantitative Analysis.

Compounds with a peak intensity sample:blank ratio < 3 were removed for each condition. Peak areas were normalized using the total ion count (TIC).

The resulting compounds were evaluated through a Student's *t*-test student ( $P < 0.05$ ) analysis to identify differentially accumulated compounds between the control samples and the MYB-targeted samples. The spectrum profile of the differential compounds is included in Supplementary Table 17.

## **Liquid chromatography and untargeted analysis**

The material used for the GC-MS analysis was further analysed by LC-MS analysis. First symptomatic leaf samples of five different plants infected with each PVX\_VIGR vector were collected at 14 dpi. Frozen ground tissue (50 mg) was extracted in 75% acetonitrile in water (500  $\mu$ L) with 1 ppm genistein as the internal standard. The homogenate was vortexed, sonicated for 10 min, and centrifuged for 5 min at 14,000 rpm. Supernatants were filtered with a 0.2  $\mu$ M filter. The analysis comprised five biological replicates per genotype.

The LC-MS analysis was performed with an Orbitrap Exploris 120 mass spectrometer coupled with a Vanquish UHPLC System (Thermo Fisher Scientific, Waltham, MA, USA). Compounds were separated by reverse-phase ultraperformance liquid chromatography using an Acquity PREMIER BEH C18 UPLC column (1.7  $\mu$ M particle size, dimensions 2.1 x 150 mm) (Waters Corp., Milford, MA, USA).

Mobile phases consisted of 0.1% formic acid in water (phase A) and 0.1% formic acid in acetonitrile (phase B). The solvent gradient program was conditioned as follows: 0.5% solvent B over the first 2 min, 0.5–30% solvent B over 25 min, 30–100% solvent B over 13 min, 2 min at 100% B, return to the initial 0.5% solvent B over 1 min, and conditioning at 0.5% B for 2 min. The flow rate was 0.4 ml/min, and the injection volume was 1  $\mu$ L. The column temperature was set at 40°C.

Ionisation was performed with heated electrospray ionization (H-ESI) in positive and negative modes. Samples were acquired in full scan mode (resolution set at 120000 measured at FWHM) and mixes were acquired in both full scan and data-dependent

acquisition (DDA) to help with compound identification. For DDA, the resolution was set at 30000 and the intensity threshold at  $2e5$ . The mass range was set from 150 to 1500.

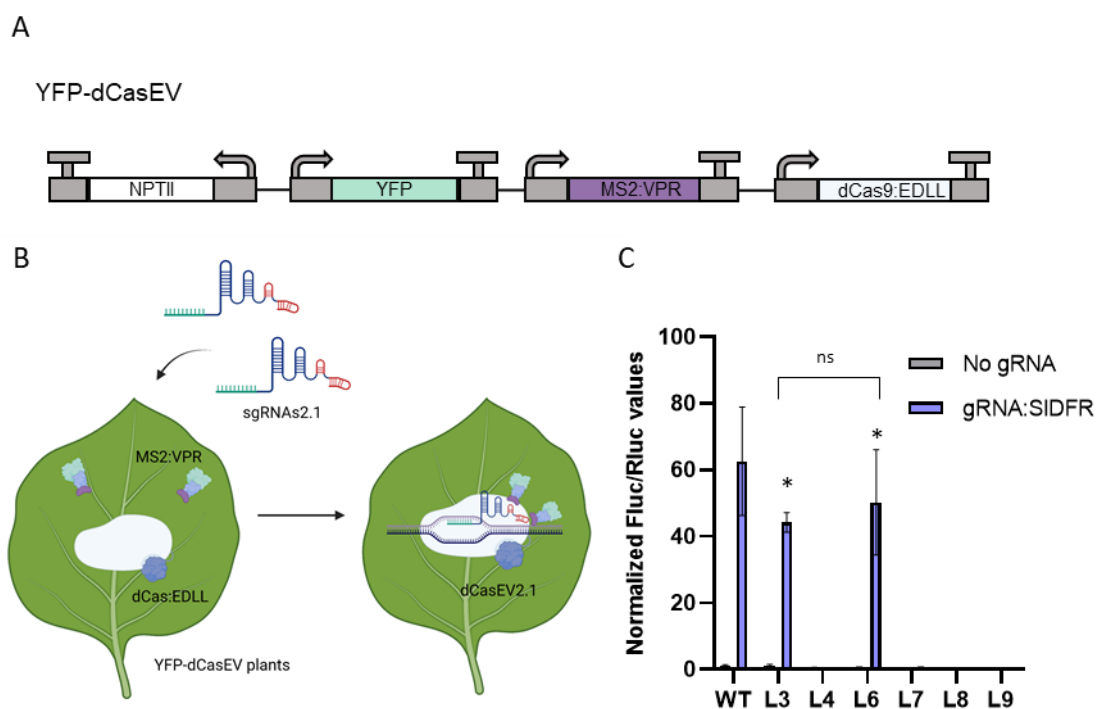
Data analysis and statistics were performed with Compound Discoverer 3.3 software (Thermo Scientific, Waltham, MA, USA) and the MetaboAnalyst5.0 Software (<https://www.metaboanalyst.ca/>). Logarithmic transformation and autoscaling scaling were employed as normalization to perform the principal component analysis and the hierarchical clustering and to produce a heatmap. Euclidean distance and Ward clustering algorithm were applied as parameters for making hierarchical clustering, and the ANOVA test was the statistical method used for generating the list of 350 molecular features with significantly different accumulation. The molecular features determined as significantly different and tentatively identified through Kegg databases associated with the Compound Discoverer 3.3 software were manually added to Table 1. The quantification of the metabolites was performed employing the parental ion and normalized with the negative control samples (C-), where an unspecific non-specific gRNA was delivered with the PVX\_VIGR.

## RESULTS

### Generation of dCasEV2.1 *N. benthamiana* chassis for CRISPRa

The reported ability of CRISPR-based transcriptional activators to efficiently regulate gene expression in *N. benthamiana* prompted us to generate a transgenic line to serve as the chassis for CRISPR-mediated targeted transcriptional activation (CRISPRa) experiments. To achieve this, a construct carrying the constant genetic elements in the dCasEV2.1 strategy, along with a yellow fluorescent protein (YFP) (Figure 1A), was employed to generate the new transgenic lines, named YFP-dCasEV. Upon transient delivery to the plant cells of one or more gRNAs targeting a specific promoter, a full dCasEV2.1 ribonucleoprotein complex will be assembled, which would expectedly result in a strong and specific gene activation (Figure 1B). The YFP-positive transgenic plants obtained in *Agrobacterium*-mediated transformation were then evaluated in a CRISPRa assay where a gRNA targeting the tomato *dihydroflavonol 4-reductase* (*SIDFR*) promoter

was delivered via agroinfiltration. Targeted activation was scored using a luciferase reporter driven by the *SIDFR* promoter and co-delivered together with the gRNA. Six transgenic lines were generated, but only two of them, namely YFP-dCasEV\_L3 and YFP-dCasEV\_L6, showed a strong gene activation upon specific gRNA expression and no activation in the absence of gRNA. As a positive control, a wild type *N. benthamiana* line was transiently co-infiltrated with the dCas9EV2.1 module, gRNA, and the luciferase reporter. No statistically significant differences were found between the two positives, stably transformed lines and the transiently infiltrated control (Figure 1C). We decided to establish YFP-dCasEV\_L6 as our chassis line, from which we later obtained a homozygous T2 population.



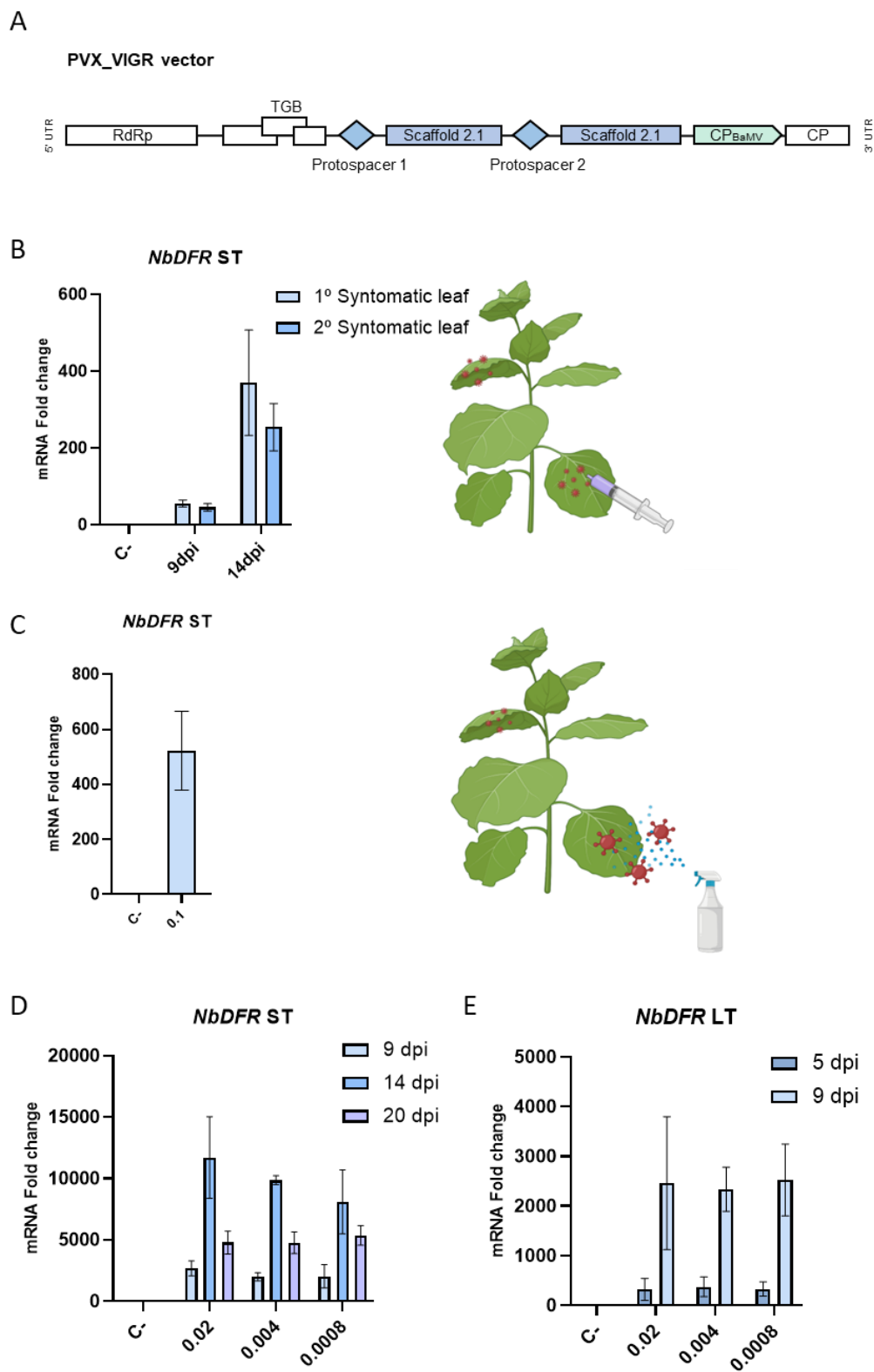
**Figure 1: Generation and selection of *N. benthamiana* chassis for CRISPRa.** A) Representation of the DNA construct employed for *N. benthamiana* transformation that includes the constant module of the dCas9EV2.1 (dCas9:EDLL and MS2:VPR), a YFP reporter (YFP), and a Kanamycin resistance (*NPTII*) gene. B) Schematic representation of action mechanism of YFP-dCasEV plants by the exogenous incorporation of gRNAs (gRNA2.1), forming the dCasEV2.1 complex as a transcriptional activation system. C) Normalized Fluc/ Rluc ratios measured in transgenic lines containing the constant module

of dCasEV2.1. The right side of the leaf transiently expresses a luciferase reporter driven by the *SIDFR* promoter and a gRNA that targets the position -150 relative to the TSS of the *SIDFR* promoter (gRNA:SIDFR). The left side of the leaf transiently expresses a luciferase reporter driven by the *SIDFR* promoter as a negative control of induction (no gRNA). WT represents a non-transgenic plant that transiently expresses the dCasEV2.1 components. Asterisks indicate Student's t-test significant values ( $P < 0.05$ ) compared with the control samples. L3, L4, L6, L7, L8, and L9 are independent YFP-dCasEV T1 lines. Bars represent average relative transcriptional activities (RTAs)  $\pm$  SD,  $n = 3$ . Images generated with BioRender.com

## Engineering PVX-based sgRNA delivery for systemic gene activation

Once the CRISPRa chassis was developed, it was necessary to optimize the gRNA delivery for gene activation. Taking the previously reported work by Uranga et al. (2021a) as a starting point, a new version of the pPVX-based gRNA delivery vector was designed (PVX\_VIGR vector), changing the native gRNA scaffold for the so-called gRNA2.1 scaffold, which incorporates two RNA aptamers that recognize the MS2 phage coat protein at the 3' end of the scaffold. Also, a multiplexing structure harbouring two tandemly arrayed gRNAs under the control of the PVX CP promoter was incorporated into the recombinant pPVX viral vector (Figure 2A). Next, the efficiency of PVX\_VIGR as an agroinfection-mediated delivery agent for gRNAs in systemic gene activation was tested, using the CRISPRa-ready dCasEV\_L6 *N. benthamiana* line as recipient chassis. For this assay, we targeted the endogenous *NbDFR* gene, whose strong and highly specific responsiveness to dCasEV2.1 activation was earlier reported (Selma *et al.*, 2019). Two *NbDFR*-specific gRNAs were designed for this purpose, targeting the positions -145 and -198 bp relative to the TSS of the *NbDFR* gene, and cloned into the PVX\_VIGR vector, generating the PVX::gDFR construct. In a first approach, the PVX\_VIGR-encoded gene activation program was agroinoculated in basal *N. benthamiana* leaves, employing an optical density (OD) at 600 nm of 0.5 as was described previously in Uranga et al 2021a. The activation effects were recorded in the first and second symptomatic leaves, 9 and 14 days post-inoculation (dpi), respectively. The symptoms are easily noticeable due to the

characteristic phenotypic alterations generated in the leaves. As shown in Figure 2B, strong transcriptional activation of *NbDFR* was observed in symptomatic leaves in plants treated with PVX::gDFR construct, but not in those treated with a PVX\_VIGR vector carrying control gRNA. No significant differences in the activation rates were found between the two analysed symptomatic leaves, but a considerable increase in transcript levels was observed over time, reaching activation levels of 45- to 55-fold at 9 dpi and 250- to 370-fold at 14 dpi. The assay was repeated including a 20 dpi time point; however, no significant increase in the activation rates was observed in this time point (Supplementary Figure 5).



**Figure 2: Engineering and optimization of PVX-VIGR vector for gRNA delivery.** ) Schematic representation of recombinant PVX\_VIGR with the gRNA2.1. The gRNA2.1 is a modification



---

of the native gRNA scaffold with the addition of two RNA aptamers at 3' that recognize the MS2 phage coat protein. B) *NbDFR* mRNA fold change at 9 and 14 dpi measured in the first and second systemic symptomatic leaves (ST, systemic tissues), after targeting *NbDFR* at positions -145 and -198 (relative to the TSS) with PVX-VIGA by syringe agro-inoculation. C) Same as in (B) but using spray inoculation at an optical density (OD) of 0.1. D) *NbDFR* mRNA fold changes observed when PVX-VIGA was inoculated at optical densities of 0.02, 0.004, and 0.0008 through spray application. E) *NbDFR* mRNA fold change in local tissues (LT) at 5 and 9 dpi using the same ODs as in (D). The C- represents a negative control where a non-specific gRNA was delivered with the PVX\_VIGR. Bars represent average RTAs  $\pm$  SD, n = 4. Images generated with BioRender.com.

Our initial results prompted us to implement further optimizations, following the logic that the usability of VIGR in, for example, large scale applications would depend on the simplicity of the delivery method and loads of the delivery agent required. For that reason, a spray-based agro-infection was employed as an alternative method to classical agro-infiltration with a syringe or vacuum. This method has some advantages such as a non-invasive application and the possibility of infecting many plants at the same time with simple equipment. As a first approach, an *Agrobacterium* optical density (OD) at 600 nm of 0.1 was employed, spraying two leaves per plant. After 14 dpi, *NbDFR* transcript levels in systemic tissues showed a successful activation of the gene, reaching up to 520-fold in the first symptomatic leaf as compared with control plants, where non-specific gRNA was delivered with the PVX\_VIGR (Figure 2C). The next analysis was focused on obtaining a minimal effective OD for activation. In this case, also the youngest sprayed leaf was analysed at 5 and 9 dpi to measure the effects in local tissues of the PVX\_VGR recombinant virus. Interesting, no significant differences in gene activation were found among the optical densities of 0.02, 0.004 and 0.0008, both in systemic (Figure 2D) or local infections (Figure 2E). Surprisingly, the lowest optical densities assayed in this second experiment resulted in activation rates even higher than those obtained in assays at OD= 0.1, reaching up 11,000-fold activation at 14 dpi in the first upper symptomatic leaf. As in the previous experiment, no further increases in the activation rates were observed at 20dpi. On the contrary, a slight decrease in the *NbDFR* transcript levels was observed at the later time point, perhaps due to leaf ageing. Also, successful activation of *NbDFR* was achieved in local leaves by the spraying method, with

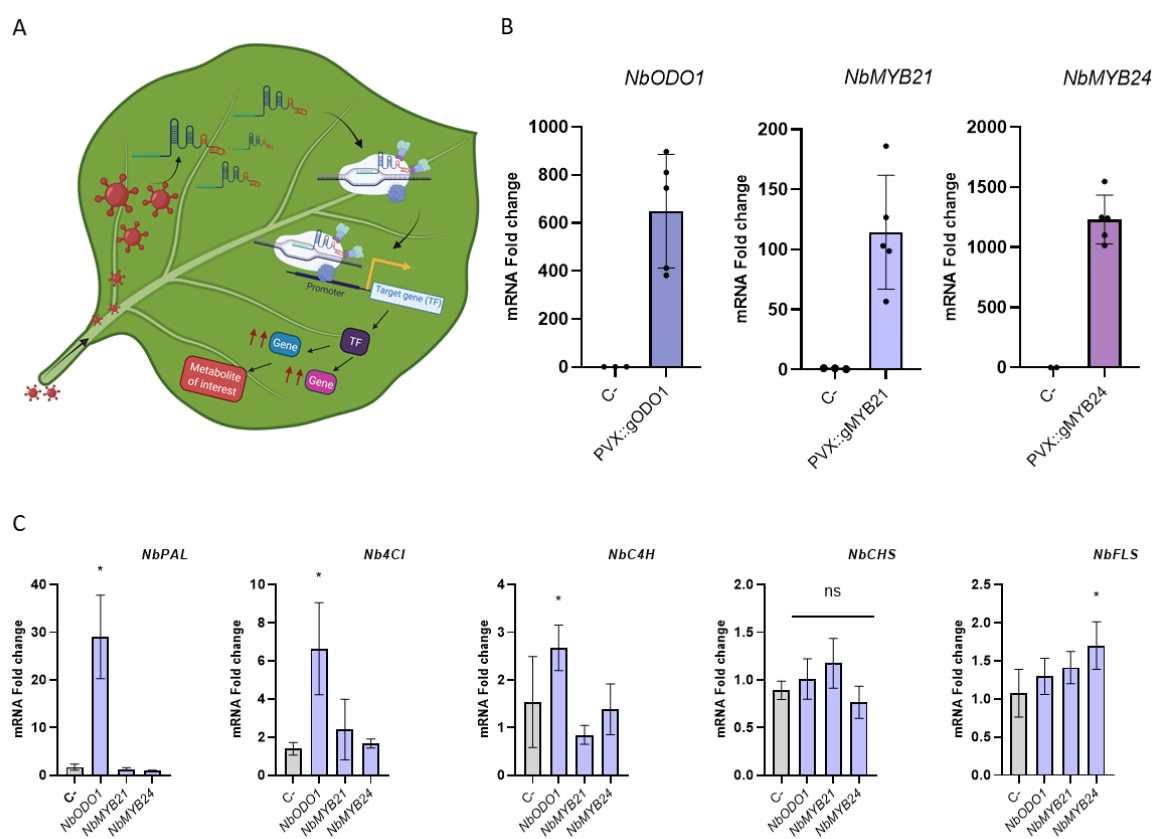
maximum activation rates of 2500 fold found at 9 dpi. These results confirm that the spray approach is an effective, easy-to-handle, non-invasive method to deliver to the plant the activation information contained in the PVX-VGR system.

### **PVX\_VGR spray activates three different *N. benthamiana* MYB transcription factors generating distinctive transcriptional responses**

To evaluate the ability of PVX\_VGR to activate endogenous genes other than *NbDFR*, we designed specific activation gRNAs programs for three endogenous MYB transcriptional factors (TFs), namely *NbODO1* (NbD050495.1), *NbMYB21* (NbD050797.1) and *NbMYB24* (NbD027779.1). The criteria for the selection of these three factors were that they (i) show high levels of expression in at least one tissue (flowers) but low levels in the leaves, and (ii) show homology with MYB factors from other species involved in the regulation of secondary metabolic profiles. With the first criterium, we wanted to ensure the inducibility of the selected genes, avoiding targeting pseudogenes or constitutively inactivated gene homeologues. With the second criterium, we aimed to show the possibility to create distinctive MYB-specific metabolite fingerprints in leaves using PVX\_VGR (Figure 3A). Highly homologous genes for all three TFs were earlier described as modifiers of the phenylpropanoid composition and/or of the volatile profiles in *Arabidopsis*, *petunia* and *tomato*, respectively (Verdonk *et al.*, 2005; Dal Cin *et al.*, 2011; Huang *et al.*, 2020). Very little is known, however, about their role in *N. benthamiana*.

Two gRNAs were designed for activating each of the target genes. The criteria for protospacer selection were: (i) location between -100 and -300 bp upstream of the TSS (Pan *et al.*, 2021a), (ii) optimal on-target score, and (iii) absence of putative off-targets in the *N. benthamiana* genome. Following the same multiplexing strategy described above, the gRNAs were incorporated in the PVX\_VGR vector and delivered with the spray-based method using an OD (600 nm) of 0.0008. This concentration was chosen since no differences were found after comparing the activation rates achieved with more concentrated PVX\_VGR *Agrobacterium* suspensions. Next, the transcriptional activation of the targeted TFs was analysed at 14dpi in the first symptomatic leaves by

RT-qPCR. The results in Figure 3B indicate that all three PVX-delivered programs resulted in significant activation of the target genes, which are otherwise repressed in leaves, reaching activation rates of >600-fold for *NbODO1*, >110 fold for *NbMYB21*, and >1200-fold for *NbMYB24* transcripts, respectively. As an indirect indication of the absolute mRNA levels obtained, the  $C_t$  values of all three TFs exceeded those observed for the *F-box* gene employed as a normalizing reference (between 25-26  $C_t$ , Supplemental Table 4). Next, we analysed the changes in the transcript levels of putative downstream-regulated genes in the phenylpropanoid pathway: *NbPAL*, *NbCHS*, *Nb4CL*, *NbC4H*, and *NbFLS* (Battat *et al.*, 2019; Boersma *et al.*, 2021). As shown in Figure 3C, only the *NbODO1* activation program resulted in a strong upregulation of the *NbPAL*, *Nb4CL*, and *NbC4H* genes, whereas the *NbMYB24* activation program only generated a slight increase in *NbFLS*, and *NbMYB21* did not result in significant activation in any of the analysed genes.

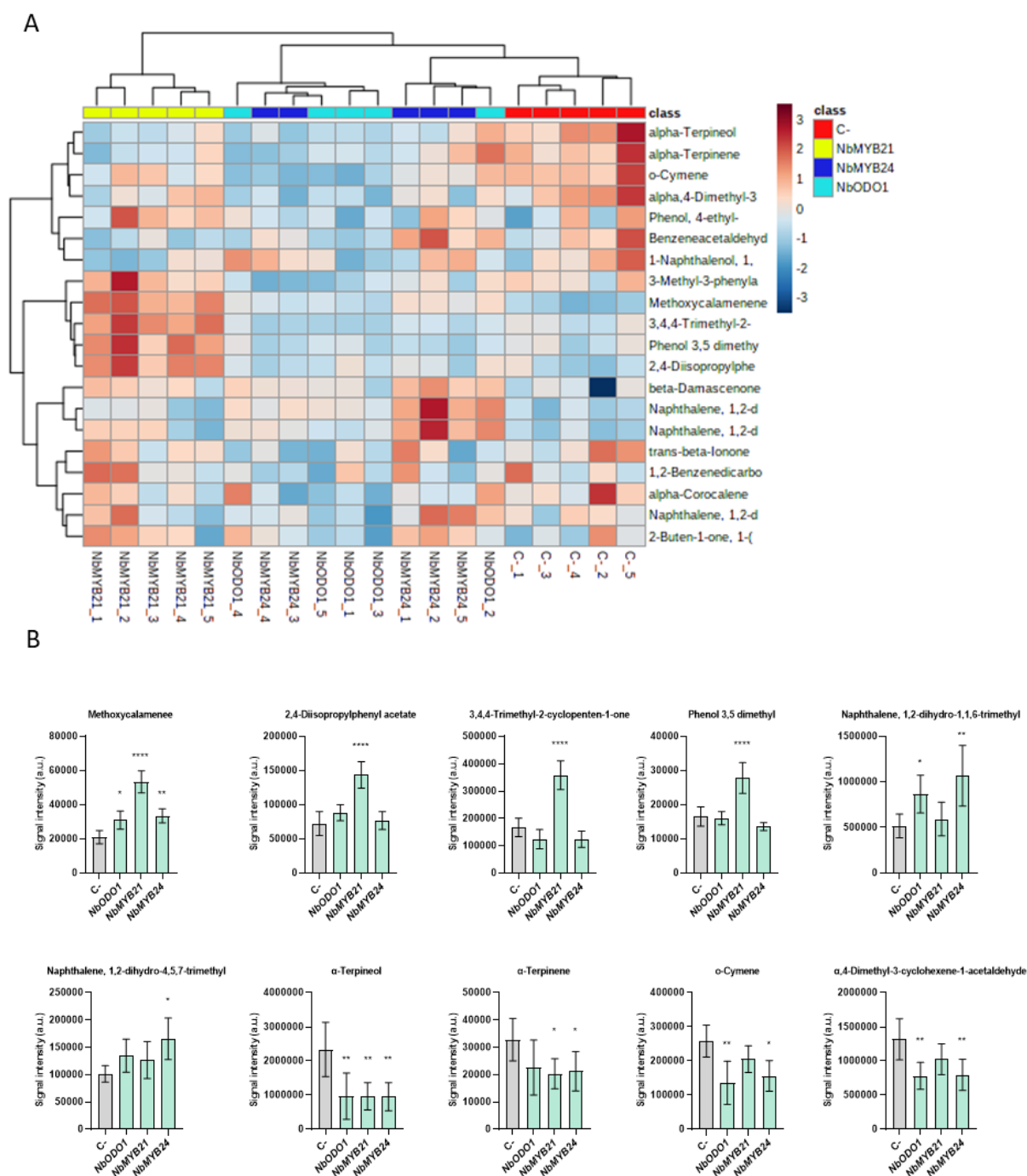


**Figure 3: PVX\_VIGR gRNA delivery for systemic activation of TFs** A) Schematic representation of the PVX\_VIGR vector employed for altering the metabolite profiles. B) mRNA fold change at 14 dpi in the first symptomatic leaf (ST, systemic tissue) upon targeting

the TFs *NbODO1*, *NbMYB21*, and *NbMYB24* with dCasEV2.1 by spraying the PVX\_VIGR vectors at an optical density (OD) of 0.008. C) mRNA fold change of *NbPAL*, *NbC4H*, *Nb4Cl*, *NbCHS*, and *NbFLS* genes at 14 dpi in the first symptomatic leaf in plants where the *NbODO1*, *NbMYB21*, or *NbMYB24* were activated with dCasEV2.1 employing a spray inoculation of PVX\_VIGR vectors at an OD of 0.008. The C- represents negative controls where a non-specific gRNA was delivered with the PVX\_VIGR. Asterisks indicate Student's *t*-test significant values ( $P < 0.05$ ) compared with the control samples. Bars represent average RTAs  $\pm$  SD,  $n = 5$  in TFs induction analysis and  $n = 3$  in *NbPAL*, *NbC4H*, *Nb4Cl*, *NbCHS*, and *NbFLS* transcriptomic analysis. Images generated with BioRender.com.

### **PVX\_VGR sprayed leaves show target-specific metabolic profiles**

Once the distinctive transcriptomic responses produced by each PVX\_VIGR activation program were demonstrated, we decided to investigate if this had a reflection on the leaf metabolite composition, yielding three distinctive metabolite profiles. For this, we first analysed the volatile profiles of PVX\_VGR-treated leaves. It is shown in Figure 4A, although no drastic differences were found in the volatile profiles when the 20 most significantly different volatiles were displayed in a hierarchical cluster. In general, a significant decrease of some monoterpenes such as alpha-terpineol was observed in all the MYB-activated samples compared with PVX\_VGR-treated control samples carrying gRNAs with no match in the genome. In upregulation terms, the greatest changes were found in the *NbMYB21*-sprayed plants. Here methoxycalamenene (a sesquiterpene) showed a 2,5-fold increase, and other compounds with unknown biological functions, such as 3,4,4-trimethyl-2-cyclopentene-1-one, 2,4-diisopropylphenyl acetate, and phenol 3,5-dimethyl, showed *NbMYB21*-specific overaccumulation (Figure 4B).

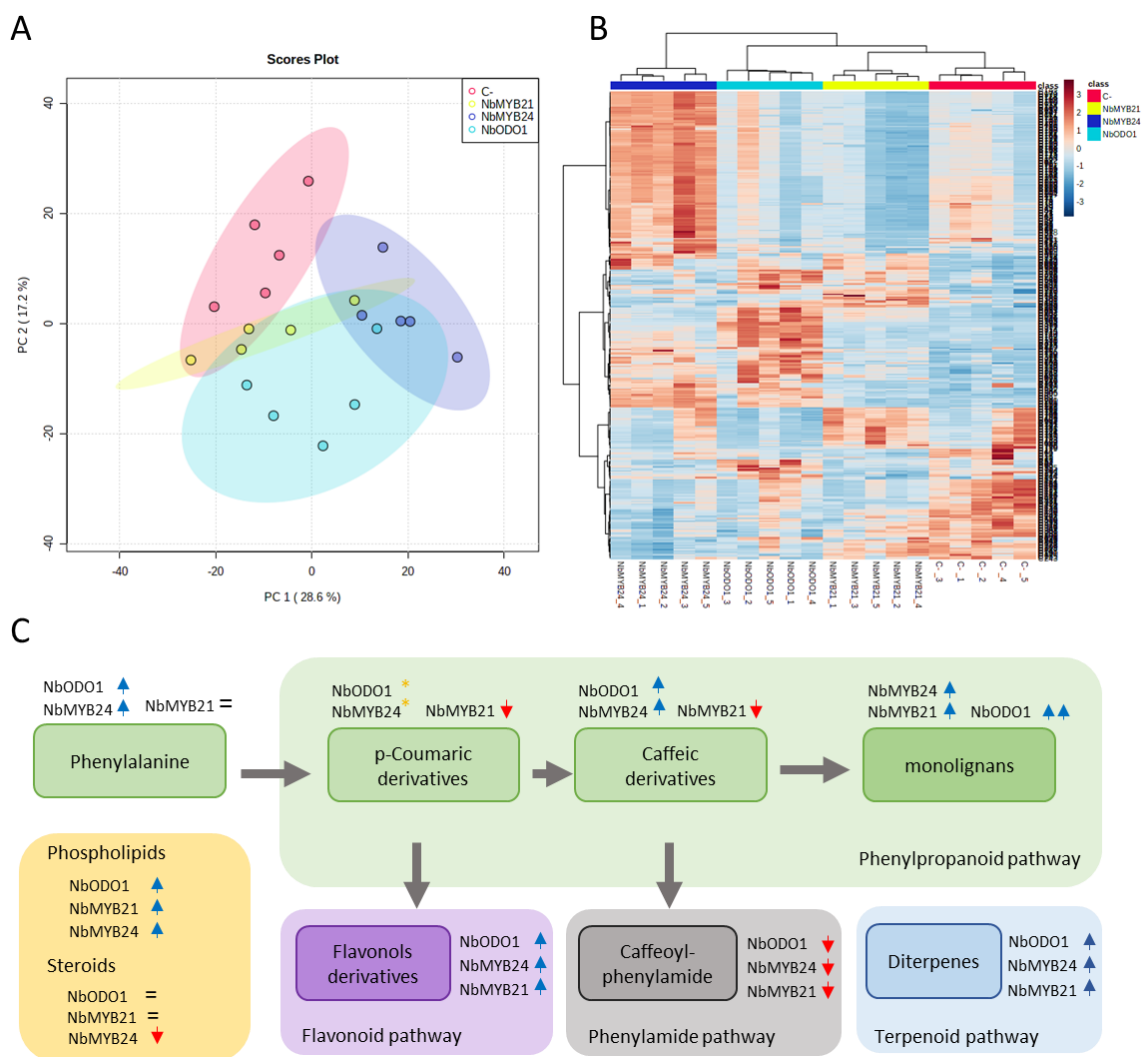


**Figure 4: Analysis of the volatile profiles of *N. benthamiana* leaves activated in *NbODO1*, *NbMYB21* and *NbMYB24* expression.** A) Hierarchical cluster analysis and heatmap representation of the volatile metabolite profiles obtained by GC-MS from leaves treated with PVX\_VIGR vectors targeting *NbODO1*, *NbMYB21* and *NbMYB24*. Control samples (C-) were treated with a PVX\_VIGR vector carrying a non-specific gRNA. The cluster shows the 20 most significantly different metabolites using a Student's *t*-test analysis ( $p < 0.05$ ). The data were obtained using Euclidean distance and Ward's minimum variance method. Red indicates up-regulated, and blue indicates down-regulated metabolites. Five biological

replicates were employed per condition. B) Signal intensity analysis of the differential compounds found in at least one of the PVX\_VIGR treatments ( $P < 0.05$ ). Bars represent average fold change  $\pm$  SD ( $n = 5$ ). Asterisks indicate Student's t-test significant values (\* =  $p < 0.05$ , \*\* =  $p < 0.01$  and \*\*\*\* =  $p < 0.0001$ ).

The modest differences observed in the non-targeted analysis of the volatile compounds prompted us to evaluate the non-volatile metabolic profiles. For that, an untargeted LC-MS analysis was carried out with the same samples previously analysed by RT-qPCR and GC-MS. Interestingly, LC-MS revealed treatment-specific metabolic fingerprints, as the principal component analysis (PCA) in Figure 5A shows. The class-specific differences are depicted in Figure 5B where the 350 most significantly different features are hierarchically clustered. The transcription factor that generates the most distinctive non-volatile metabolic fingerprint is *NbMYB24*, followed by *NbODO1*. In contrast, *NbMYB21* activation shows a lower ability to alter the metabolite content of *N. benthamiana* leaves. A tentative identification of differential compounds, performed by the differential  $m/z$  ions and their respective retention times, was carried out as listed in Table 1 and summarized in Figure 5C. As expected, most perturbations corresponded to metabolites in the phenylpropanoid pathway. This is reflected in the increased levels of phenylalanine in *NbODO1*- (1.6 fold) and *NbMYB24*- (1.9 fold) induced samples, and in the increased levels of coumaroyl-*D*-quinic acid, the product of the *PAL*, *C4H* and *CL4* gene activities, in the samples sprayed with *NbODO1* (2.3fold) and *NbMYB24* (2.7 fold) activation programs. The accumulation of 5-O-caffeoyl shikimic acid and caffeic acid 3-glucoside in the *NbMYB24* (5.1and 2.2-fold respectively) and *NbODO1* samples (4.4 and 2.3-fold respectively) indicates that the activation of these two TFs leads to the upregulation of the caffeic acid branch of the phenylpropanoid pathway. On the contrary, a clear downregulation of chlorogenic acid is observed in *NbMYB21* samples (0.3-fold). Furthermore, a general reduction in N-caffeoylputrescine was observed in all three TF activated samples. Finally, in *NbODO1* samples, the metabolic flux of phenylpropanoid derivatives presents a strong shift towards lignans accumulation, as evidenced by the high accumulation of syringaresinol  $\beta$ -D-glucoside (> 22-fold) and sinapoylglucose (> 4-fold). Slight increases in these compounds can also be observed in

*NbMYB24* and *NbMYB21* samples but to lower extent. Remarkably, flavonols over-accumulate in all TF-targeted samples, specifically kaempferol glucosides, with the highest accumulation observed in *NbODO1* samples (11-fold). Finally, other metabolic pathways also seem perturbed, although to a lower extent. A significant increase in phospholipids (Linoleoylglycerol and LysoPC), and a decrease in hydroxy steroid derivatives (ethylestrenol) were found in *NbMYB24*-induced samples.



**Figure 5: Analysis of the metabolite profiles of *N. benthamiana* plants sprayed with PVX\_VIGR for targeting *NbODO1*, *NbMYB21* and *NbMYB24*.** A) Principal component analysis and (B) Hierarchical cluster analysis and heatmap representation resulting from the non-targeted LC-MS data obtained from leaves treated with PVX\_VIGR targeting *NbODO1*, *NbMYB21*, or *NbMYB24*. Control samples (C-) were treated with a PVX\_VIGR vector carrying a non-specific gRNA. Five biological samples were employed per condition. The *m/z* features represented in the heatmap are the 350 most significantly different using an ANOVA test ( $P < 0.05$ ). The data were obtained using Euclidean distance and Ward's minimum variance method. Red indicates up-regulated and blue down-regulated features. C) Schematic representation of the metabolic pathways and metabolite sub-groups that show differential abundance after inducing the *N. benthamiana* plants with a PVX-VIGR for targeting *NbODO1*, *NbMYB21*, or *NbMYB24* TFs compared with control samples (C-). The blue arrows represent an increment of the indicated metabolite group in the samples where *NbODO1*, *NbMYB21* or *NbMYB24* were up-regulated with dCasEV2.1. The red arrows represent a down-regulation of the indicated metabolite group in the samples where *NbODO1*, *NbMYB21*, or *NbMYB24* were up-regulated with dCasEV2.1. The orange asterisk represents that some metabolites that belong to the indicated chemical class are increased and other metabolites are down-regulated in the samples where *NbODO1*, *NbMYB21*, or *NbMYB24* were up-regulated with dCasEV2.1. The = symbol represents no differential metabolite changes in the sample where one of the indicated MYB TF was up-regulated with dCasEV2.1. The metabolites employed to generate this illustration were extracted from Table 1 where the differential metabolites tentatively identified are included.



**Table 1. Differential tentatively identified compounds from Non-targeted LC-MS analysis of the *NbODO1*, *NbMYB21* and *NbMYB24* pPVX induced samples, normalized with the negative control (C-) values. The colouration of the cells for each compound corresponds to the colours employed in the diagram of the metabolic routes in Figure 5C. Asterisks indicate T student significant values (\* =  $p < 0.05$ , \*\* =  $p < 0.01$  and \*\*\*\* =  $p < 0.0001$ ).**

Compounds	Mass	RT (min)	Fold Change			
			C-	NbODO1	NbMYB21	NbMYB24
L-Phenylalanine	165,08	4,81	1,00 ±0,01	1,60 ±0,45*	1,36 ±0,17	1,90 ±0,09****
N-Caffeoylputrescine I	250,13	5,04	1,00 ±0,62	0,33 ±0,19**	0,12 ±0,10****	0,31 ±0,21*
N-Caffeoylputrescine II	250,13	5,53	1,00 ±0,64	0,30 ±0,15**	0,12 ±0,10****	0,30 ±0,13*
N-Caffeoylputrescine III	250,13	7,10	1,00 ±0,53	0,33 ±0,19	0,13 ±0,09****	0,35 ±0,22
Chlorogenic acid I	354,09	7,39	1,00 ±0,28	0,43 ±0,16	0,29 ±0,22**	1,03 ±0,32
Kaempferol-3-O-glucuronide	462,10	7,59	1,00 ±0,47	11,12 ±3,13****	5,53±2,23****	9,94 ±9,07****
Chlorogenic acid II	354,09	7,63	1,00 ±0,26	0,76 ±0,21	0,27 ±0,25**	1,49 ±0,42
Chlorogenic acid III	354,09	7,80	1,00 ±0,23	0,81 ±0,22	0,27 ±0,25**	1,59 ±0,45
Benzenoid derivative	659,16	7,99	1,00 ±0,52	0,06 ±0,09****	0,36 ±0,20	0,49 ±0,10
5-O-Coumaroyl-D-quinic acid	338,10	9,50	1,00 ±0,12	1,30 ±0,28	0,79 ±0,16	2,28 ±0,29****
Chlorogenic acid IV	354,09	10,17	1,00 ±0,21	0,97 ±0,38	0,42 ±0,31	2,02 ±0,26*
Caffeyl alcohol derivative	166,06	10,58	1,00 ±0,09	2,66 ±0,40****	1,08 ±0,07	1,49 ±0,28**
Caffeic acid 3-glucoside	342,09	10,65	1,00 ±0,17	2,29 ±0,86****	1,52 ±0,15	2,24 ±0,17**
Chlorogenic acid V	354,10	10,87	1,00 ±0,23	0,96 ±0,34	0,40 ±0,31*	1,95 ±0,10**

Table 1 (continued).

Compounds	Mass	RT (min)	Fold Change			
			C-	NbODO1	NbMYB21	NbMYB24
5-O-Coumaroyl-D-quinic acid	338,10	11,82	1,00 ±0,13	1,98 ±0,34***	1,24 ±0,18	2,36 ±0,13***
5-O-Coumaroyl-D-quinic acid	338,10	12,31	1,00 ±0,16	1,84 ±0,38**	0,99 ±0,25	2,38 ±0,19***
Chlorogenic acid IV	354,09	12,38	1,00 ±0,24	0,70 ±0,27	0,41 ±0,25*	1,47 ±0,25
5-O-Coumaroyl-D-quinic acid	338,10	12,71	1,00 ±0,08	2,34 ±0,39***	1,04 ±0,26	2,71 ±0,15***
Kaempferol 3-O-gentiobioside-7-O-rhamnoside	756,21	12,98	1,00 ±0,13	2,86 ±0,58***	1,52 ±0,29*	2,03 ±0,59***
5-O-Caffeoylshikimic acid	336,08	13,00	1,00 ±0,44	4,39 ±2,22***	0,68 ±0,49	5,10 ±0,91***
Kaempferol 3-O-gentiobioside-7-O-rhamnoside	756,21	13,09	1,00 ±0,34	2,57 ±0,72***	1,73 ±0,25*	2,27 ±0,48***
Sinapoylglucose (lignan)	386,12	13,27	1,00 ±0,17	4,44 ±0,94***	1,56 ±0,25*	2,38 ±0,12***
N-Benzoylglutamic acid	251,08	13,67	1,00 ±0,22	1,91±0,4***	1,38 ±0,13*	1,64 ±0,09**
5-O-Coumaroyl-D-quinic acid	338,10	14,47	1,00 ±0,14	1,60 ±0,3**	0,95 ±0,23	1,97 ±0,09***
Lignan glucoside (Isolariciresinol 9'-O-beta-D-glucoside)	764,42	17,25	1,00 ±0,14	3,01 ±0,64**	1,40 ±0,23	2,13 ±0,33***
Kaempferol-rutinoside II	594,16	18,25	1,00 ±0,24	2,53 ±0,76***	1,84 ±0,19*	2,38 ±0,61***
Coumaryl-hexose malic acid I	442,18	19,05	1,00 ±0,24	0,28 ±0,34	0,33 ±0,04	0,38 ±0,15*
Syringaresinol β-D-glucoside	580,22	19,73	1,00 ±0,14	<b>22,96</b> <b>±3,56***</b>	1,84 ±0,28***	3,49 ±0,36***
Dihydroxy benzoic acid derivative I	420,11	23,03	1,00 ±0,12	0,33 ±0,09***	0,57 ±0,05***	0,39 ±0,03***
Coumarolyquinic acid glucoside derivative	1041,50	26,96	1,00 ±0,24	0,52 ±0,35	0,33 ±0,20*	0,29 ±0,18*

Table 1 (continued).

Compounds	Mass	RT (min)	Fold Change			
			C-	NbODO1	NbMYB21	NbMYB24
Coumarolyquinic acid glucoside derivative	1041,50	27,66	1,00 ±0,29	0,46 ±0,34	0,28 ±0,18*	0,30 ±0,17*
Coumarolyquinic acid glucoside derivative	1041,50	28,05	1,00 ±0,36	0,30 ±0,29	0,22 ±0,16*	0,18 ±0,18*
Coumarolyquinic acid glucoside derivative	1041,50	28,18	1,00 ±0,31	0,48 ±0,29	0,25 ±0,16*	0,30 ±0,15*
Coumarolyquinic acid glucoside derivative	1041,50	28,67	1,00 ±0,30	0,54 ±0,30	0,27 ±0,16*	0,28 ±0,17*
Coumarolyquinic acid glucoside derivative	1041,50	28,76	1,00 ±0,31	0,43 ±0,34	0,33 ±0,17*	0,26 ±0,16*
Coumarolyquinic acid glucoside derivative	1041,50	28,93	1,00 ±0,23	0,42 ±0,32	0,41 ±0,20*	0,32 ±0,14*
Hydroxy steroid derivative (ethylestrenol)	288,25	30,55	1,00 ±0,41	0,31 ±0,33	0,21 ±0,20*	0,14 ±0,14*
Hydroxy steroid derivative (ethylestrenol)	288,25	30,70	1,00 ±0,44	0,27 ±0,35	0,20 ±0,17*	0,13 ±0,12*
Linolenoylglycerol (monolinolenin)	352,26	33,84	1,00 ±0,38	2,01 ±0,36**	2,42 ±0,36***	1,95 ±0,63**
Lysophospholipid (LysoPC)	517,32	34,01	1,00 ±0,11	3,56 ±0,94***	1,95 ±0,17**	1,47 ±0,19**
Linoleoylglycerol (1-Linoleoyl-2-Hydroxy-sn-glycero-3-PC)	519,33	34,89	1,00 ±0,14	2,78 ±0,97**	2,15 ±0,13**	1,25 ±0,24
Linolenoylglycerol (monolinolenin)	352,26	34,95	1,00 ±0,18	1,59 ±0,44**	2,01 ±0,39**	2,18 ±0,95**
Diterpenoid (Trigonosin C)	608,26	38,47	1,00 ±1,01	6,71 ±4,29**	8,39 ±1,10**	8,81 ±0,97**
Diterpenoid (Trigonosin C)	608,26	38,87	1,00 ±1,01	7,87 ±4,55**	8,63 ±1,20**	8,07 ±0,89**

## DISCUSSION

Viral systems have been widely used in plant sciences for a variety of applications, from recombinant protein production to reverse genetics. Virus-induced gene silencing (VIGS) has been extensively used for interrogating gene function in diverse species and as a shortcut to gene knockouts bypassing stable transformation. In VIGS strategies, viral clones incorporate sequences matching endogenous genes that, consequently, induce post-transcriptional gene silencing in agro-infected plants. More recently, virus-induced gene editing (VIGE) has emerged as an alternative to transgenics for efficient gene editing. In this latter case, the gene edition machinery is usually split into two parts, a constant module comprising the Cas enzyme and a variable module comprising the gRNA. The gRNA module is encoded in the viral vector whereas the constant module, larger in size, can be either stably integrated into the plant genome (Uranga *et al.*, 2021a) or delivered in a separate viral vector accepting larger cargo (Uranga *et al.*, 2021b). Different viruses have been adapted for both VIGS and VIGE strategies. Whereas for VIGS, the TRV seems the ideal choice given its wide host range, mild silencing suppression, and consequently, strong gene silencing effect, in VIGE other factors, need to be taken into consideration, such as the gRNA accumulation levels, the ability to self-process multiplexed gRNA arrays, and the cargo capacity, all of which also play a role in editing efficiency. Therefore, other viruses such as PVX or TMV present advantageous characteristics for VIGE (Cody *et al.*, 2017; Ariga *et al.*, 2020).

Similar to VIGS and VIGE, VIGR represents a new promising transient tool for plant biotechnology, with the ability not only to suppress but also to strongly induce endogenous gene expression. Thus, VIGR offers a panoply of new applications that range from gene discovery to transient genetic reprogramming of crops at a large scale, as recently proposed by Torti *et al.* (2021) for PVX and TMV-based transgene expression. VipariNama, a VIGR strategy based on TRV, was pioneer in demonstrating the possibilities of virus delivery in CRISPR-based transcriptional regulation in plants (Khakhar *et al.*, 2021). Despite the elegant display of alternative VipariNama configurations, which includes both activation and repression strategies, actual gene activation/repression levels observed with VipariNama tools were relatively modest

(between 2 and 5-fold), probably due to a limited intrinsic regulatory ability of the combination of TRV-encoded gRNA and Cas9-based synthetic transcription factors employed. Whereas these activation/repression levels were sufficient to induce measurable developmental and/or metabolic changes for some regulatory genes, there is certainly room for improvement if robust responses are required in large-scale applications. Our results here show that the PVX\_VIGR system can induce a very strong gene activation in *N. benthamiana*, both local and systemic, opening the way to new and more robust applications. In this regard, a strong VIGA effect not only serves in research to elucidate gene function but in the case of *N. benthamiana*, VIGA can also be envisioned as a tool for shaping plant metabolic content. Recently our group showed how CRISPR-based programmable transcriptional factors can be employed to channel metabolic fluxes toward the accumulation of selected metabolites in *N. benthamiana* leaves (Selma *et al.*, 2021). In that work, gRNA-encoded activation programs were delivered in the form of multiplex T-DNA constructs via agroinfiltration, a strategy that limits the scalability of the approach. This limitation could be circumvented with the advantages offered by VIGR. In particular, the spraying strategy described here offers a solution for large-scale delivery of gene reprogramming instructions to *N. benthamiana* biofactories, with possibilities to extend the technology to other crops given the broad host range of PVX. In principle, in VIGR strategies, transcriptional activation/repression will not be sustained in the next generations without re-inoculating the virus. However, the recent advances in CRISPR-based epigenetic modifications open new perspectives on the ability to induce epigenetic heritable traits using VIGR (Papikian *et al.*, 2019; Ghoshal *et al.*, 2020; Selma and Orzáez, 2021).

The PVX recombinant virus presents an architecture that allows a multiplexing strategy for expressing the gRNAs without the necessity of pre-tRNAs, as was previously demonstrated by Uranga *et al.* 2021a. The addition of activation aptamers in the gRNA scaffold that are recognized by the MS2 protein in the recombinant virus structure has not been an obstacle to expressing active gRNAs. A possible constraint of PVX\_VIGR is that cargo limitations restrict the complexity of the gene activation/repression programs that can be delivered to the plant. In metabolic reprogramming using T-DNA delivery via agroinfiltration, we previously employed up to fifteen tandemly arrayed gRNAs to

---

successfully redirect a metabolic pathway (Selma *et al.*, 2021). However, it is unlikely that such large gRNA tandem repeats could work as efficient programs in PVX\_VIGR. Future works would need to establish these limitations and circumvent them with other strategies such as the co-delivery of non-competing viruses.

The capacity of PVX\_VIGR to activate genes in *N. benthamiana* was evaluated by targeting the transcriptional factors *NbODO1*, *NbMYB21*, or *NbMYB24* and analysing the specific changes in the metabolite profiles in the induced plants. In this case, although the selected targets are predominantly expressed in flowers, their role in *N. benthamiana* remains uncharacterized, giving added value to the employment of the PVX\_VIGR to analyse the metabolite fingerprint of each TF. The MYB factors described previously as putative orthologues of our target genes were studied in petunia, Arabidopsis, and tomato. The over-expression of these genes produced changes in the volatile and non-volatile metabolic profiles in these species. In petunia, the overexpression of *PhODO1* was linked to the production of compounds involved in the scent, such as benzenoids, phenylpropanoids, fatty acids, and terpenoid derived volatiles (Verdonk *et al.*, 2005). In addition, the petunia orthologs of MYB21 and MYB24, known as EOBI and EOBI (EMISSION OF BENZENOID), respectively, are also involved in the production of the components of the scent through tight crosstalk between them and *PhODO1* (Spitzer-Rimon *et al.*, 2012). In tomato, the *ODO1* overexpression is also linked to the volatile production of benzenoids and phenylpropanoids, however, the biggest changes found in the tomato metabolome associated with *ODO1* were the increase of non-volatile phenylpropanoid derivatives, which includes some flavonoids, especially flavonols (Dal Cin *et al.*, 2011; Xie *et al.*, 2016). Concurrently, in Arabidopsis, the increase of flavonols in flowers was directly correlated with the expression of *AtMYB99* (a homologue of *ODO1*), *AtMYB21*, and *AtMYB24* (Battat *et al.*, 2019; Zhang *et al.*, 2021a). The crosstalk among all three paralogues seems to take place similarly in petunia and Arabidopsis. However, the metabolite profiles resulting from this crosstalk are different in the two species, suggesting that in each species these genes regulate the different branches of the phenylpropanoid pathway differently. In addition, in Arabidopsis, *AtMYB21* and *AtMYB24* are known to interact with genes in the jasmonate pathway, leading to an increase in the production of terpenoids (Cheng *et al.*, 2009). The

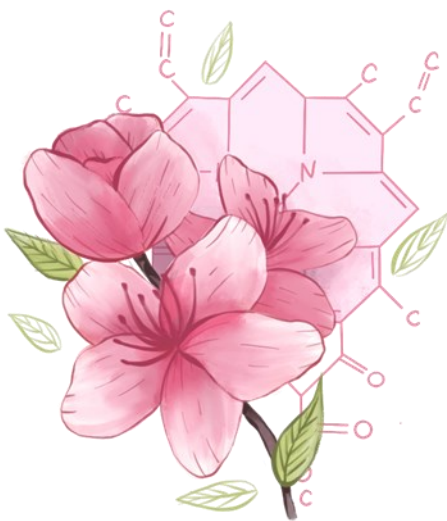
upregulation of MYB homologues in *N. benthamiana* leaves using PVX\_VIGR did not lead to dramatic changes in the volatile profiles, with the exemption of an increase in the levels of sesquiterpene methoxycalamenene observed in *AtMYB21* samples. A probable explanation is the limited capacity of *N. benthamiana* leaves to produce volatile compounds, which may be related to the high activity of endogenous glycosyl transferases producing non-volatile glycosylated derivatives (Herpen et al. 2010; Louveau and Osbourn 2019). On the contrary, highly specific non-volatile metabolic profiles were observed associated with each TF transcriptional activation. In all three treatments, most changes observed were associated with the phenylpropanoid pathway. This correlates with previous observations in *Arabidopsis* and tomato, where most changes associated with *ODO1*, *MYB21* or *MYB24* overexpression were in the non-volatile fraction, specially phenylpropanoid derivatives (Xie et al., 2016; Battat et al., 2019). In particular, we found that all three treatments resulted in the overaccumulation of flavonols. On the contrary, the accumulation of other phenylpropanoids seemed dependent on the activation program applied. For example, *p*-coumaric acid and caffeic acid derivatives were enriched in *NbMYB24* and *NbODO1* activation programs but downregulated in *NbMYB21*-induced samples. In the case of *NbODO1*, this observation correlates with the upregulation of *NbPAL* and *NbC4H* genes observed in the RT-qPCR analysis. For *NbMYB24*, we did not observe a significant upregulation of the same genes. The observed metabolite accumulation could be explained by the activation of a set of homologous genes different from those evaluated by RT-qPCR, or by a different timing in gene activation by *NbMYB24*. Other phenylpropanoid derivatives such as monolignans, although upregulated in all three treatments, resulted in much higher accumulation in *NbODO1* samples. These differential accumulations observed in certain groups of related metabolites, such as lignans or flavonols, suggest the distinctive specificities of each MYB factor in the activation of the different genes in the pathway and demonstrate the power of PVX\_VIGR in reprogramming metabolic fluxes.

In conclusion, we show that PVX-based gRNA delivery is an effective system for transient CRISPR activation. In this work, we demonstrate that plant metabolic profiles can be customized through a spray agro-inoculation of a PVX vector, expanding the current toolbox of viral systems employed for transcriptional activation. The precise and

---

scalable spatiotemporal control of gene transcription offered by this new tool opens the door to increasingly sophisticated designs in plant synthetic biology.





# General discussion



## GENERAL DISCUSSION

Synthetic Biology (SynBio) aims to redesign biological systems to obtain useful characteristics and applications, employing the engineering principles of standardization, modularity, and abstraction. The SynBio approaches build new genetic devices made of gene circuits capable of generating predictable outputs through user-designed inputs and adapting them into different chassis. The systematic design of these genetic circuits implies the connection of functional modules, usually following a general schema borrowed from other engineering sciences: sensor modules detect input signals, which are integrated and transformed by processor units and sent to final actuators, which generate output responses.

The challenge of using plants as chassis for synthetic biology for obtaining new interesting phenotypic traits and functions is to develop systematic processes of identifying and abstracting orthogonal genetic elements for building complex genetic circuits. Although examples of fully integrated genetic circuits currently exist in plants (Antunes *et al.*, 2011; Okada *et al.*, 2017), functional circuits with agronomic applications are still scarce. However, the list of available new synthetic genetic elements does not stop growing, providing sophisticated functional modules for expanding the catalogue of inputs and outputs that plants can process. The sensors represent the first step in the control of gene expression: for this reason, a large collection of these functional modules has been developed to respond to endogenous and exogenous stimuli (Leydon *et al.*, 2020). However, the responses that a sensor coupled to a processor can generate with currently available genetic tools are very limited, since most typical inducible expression systems developed for plants only operate the identity (on) and negation (off) functions. To increase the potential outputs and the complexity of the plant response, it is necessary to develop molecular tools that can process the transcriptional signals adequately. The simplicity and versatility of the emerging CRISPR tools based on a small RNA guide (gRNA) and an effector Cas protein offer an excellent approach to develop programmable transcriptional regulations thanks to the inactivated version of the Cas that binds and targets DNA sequences and to the possibility of anchoring regulation

domains to its structure (Tanenbaum *et al.*, 2014a; Konermann *et al.*, 2015b; Zalatan *et al.*, 2015; Zhang *et al.*, 2015c).

In the work presented in this thesis, an improved CRISPR-based tool that performs programmable transcriptional activations has been developed, offering an orthogonal and versatile tool for expanding the functionalities of genetic circuits in plant synthetic biology approaches.

Chapter 1 provided a detailed description of the optimizations performed over the dCas9-PTA (programmable transcriptional activators) to maximise the activation capacity of this tool, improving the dCas9 strategies for gene regulation available in plants until that moment (Piatek *et al.*, 2015; Park *et al.*, 2017; Li *et al.*, 2017b; Lowder *et al.*, 2018b). Although these early strategies were pioneers in creating programmable transcriptional activation tools in plants, offering a significant improvement of the available tools for this purpose, such as ZFN-TF or TALEs (Lindhout *et al.*, 2006; Gao *et al.*, 2014), the optimized dCasEV2.1 showed a strong activation capacity never before reported in plants. This tool was selected as the preferred dCas-PTA design after a systematic comparison of different protein architectures and combinations of TADs employing a luciferase reporter in *N. bethamiana* leaves. The reason behind this strong activation could be an additive effect of the TADs employed (EDLL and VPR), and the favourable gRNA2.1 loop, which shows an improvement in the activation rates of 20-35% compared with the original gRNA2.0 loop. The cause of this remains unclear, although the modification could favour the stability of the gRNA scaffold in the plant cell. Furthermore, an efficient multiplexing strategy that involves the activation elements required for the dCasEV2.1 tool was efficiently developed, expanding the modularity and versatility of the strategy. Nevertheless, to offer robust evidence of the activation capacity of the dCasEV2.1, several targets were evaluated showing notable but variable results in promoters such as pSIMTB and pSIDFR when evaluated using luciferase reporters. Even, the CaMV35S promoter showed significant activation, which was surprising since this is per se one of the strongest promoters available for plants and used for overexpression in different biological systems. This implies the possibility of increasing recombinant production in expression systems directed by this promoter (Ma *et al.*, 2003) by overactivation with dCasEV2.1. The evaluation of dCasEV2.1 using

reporter constructs stably integrated into the genome was also successful. Again, different activation rates were observed for the promoters *Nos*, *SIMTB* and *SIDFR*. This suggests a link between the basal expression levels of each promoter and the maximum induction rates that can be achieved with dCasEV2.1 (Haran *et al.*, 2000; Ede *et al.*, 2016). These results provide a more precise vision of what can be expected from the interaction of different types of promoters with transcriptional regulation tools in the cellular context and the maximum transcriptional levels that can be obtained, which is essential for the design of new genetic circuits.

For the final evaluation of the dCasEV2.1 CRISPRa tool, three different endogenous genes were evaluated to discern whether genes in their "unaltered" chromatin landscape could also be efficiently activated. An unprecedented activation rate was obtained targeting the endogenous *NbDFR* gene employing two different sets of three gRNAs, individually or in combination. Following the same trend, the activation rates obtained from the transcription factor gene *NbAN2* and the bHLH gene *NbAN1* were equally potent, reaching activation levels between 1000-10000-fold in the three target genes evaluated. These high transcriptional activation rates overpassing the transcriptional activation achieved when employing the endogenous TFs that naturally activate the target genes such as *NbDFR* have never been reported to date until the recent development of the CRISPR-Act3.0 strategy (Pan *et al.*, 2021c). This updated transcriptional activation tool uses a combination of the scRNA and SunTag strategies, generating a complex formed by the inactivated CRISPR protein, the gRNA2.0, the MS2 protein fused with the SunTag tail, and an ScFv antibody fused to activation domains. In addition, this strategy was efficiently transferred to other nucleases, such as dCas12b. This strategy offers improvements in transcriptional activation, especially focused on the gRNA design, since it allows the expansion of effective activation "window" where target gRNAs can be designed to maintain high activation levels. However, when analysed closely, the activation rates obtained with CRISPR-Act3.0 do not seem significantly different from those obtained with dCasEV2.1. Comparisons between the two systems performed by the authors of the CRISPR-Act3.0 strategy using the *OsBBM1* gene in rice protoplasts showed no significant activation differences when optimum targeting positions in the promoter were used. On the contrary, most of the differences

reported in favour of CRISPR-Act3.0 can be attributed to a suboptimal design of targeting sites, usually outside the window of 150 bp upstream of the TSS. Furthermore, the requirement of additional TUs to form the CRISPR-Act3.0 makes it less suitable for stable transgenics (Liu *et al.*, 2013).

Although the strong transcriptional activation capacity of this tool has been widely demonstrated in this chapter, another essential characteristic of programmable activators is target specificity. Off-target activities have been widely reported for Cas9-based editing strategies in mammals (Zhang *et al.*, 2015b), but in plants, they are less frequent (Zhang *et al.*, 2018). However, off-target effects in plant CRISPR-based PTA-regulation have not been established. For solving this, differential expression analysis was carried out to analyse in detail a transcriptome where targeted activation with dCasEV2.1 was performed. Two distinct categories of potential actuators were targeted: the *NbDFR* and *NbAN2* genes. *NbDFR* is an enzyme with unknown transcriptional regulatory roles, whereas *NbAN2* is a transcription factor that represents a regulatory node with multiple connections in the transcriptome. The data obtained from plotting the differential analysis of the *NbDFR*-affected transcriptome and the control transcriptome show the genome-wide specificity of the CRISPRa tool with an almost invariant gene expression profile, except for an *NbDFR* homologue that served as additional proof of specificity. The absence of off-targets that generate unwanted transcriptional activation can be explained because they must be located very close to the promoters. Their presence in a different location does not generate an increase in transcriptional activity. On the other hand, as expected, targeting *NbAN2* resulted in wider transcriptome changes due to the regulatory nature of this gene. This is confirmed by the presence of *NbDFR* among the pool of genes significantly induced upon *NbAN2* activation, although with much lower activation rates than in the samples where this gene was directly activated with dCasEV2.1. In conclusion, Chapter 1 shows that the improved CRISPRa tool, named dCasEV2.1, is a powerful, orthogonal, and specific programmable transcriptional activator, which makes it a versatile processor component with the capacity to be connected with different downstream functional modules (a final actuator enzyme or a master regulator). Currently, there are already some examples in the literature of dCas9-PTR tools coupled with actuators for regulating

agronomically- relevant traits such as increasing crop tolerance to drought (Roca Paixão et al. 2019; de Melo et al. 2020) or regulating the flowering time (Papikian et al. 2019; Lee et al. 2019). The dCasEV2.1 tool offers a potent strategy for developing potential agronomically interesting genetic circuits for controlling drought and temperature tolerance (Shim et al. 2018; Cheng et al. 2021), fruit ripening (Liu et al. 2022) or secondary metabolite production (Zhang et al. 2021). Furthermore, this dCas9-PTA can be efficiently connected with sensors for expanding its potential applications in the field, as was shown in the work carried out by Garcia-Perez et al. (2021) where the activating components of dCasEV2.1 are regulated by a copper sensor to offer a strict control of the CRISPR-targeted genes.

The potential applications of the dCasEV2.1 tool are not limited to control master regulators that have pre-defined downstream responses since it can also be employed for the specific activation of individual enzymatic steps to redirect the metabolic flux and increase the accumulation of a target metabolite.

In Chapter 2, a proof of concept of the modulation of the metabolic flux in plants through the application of the dCasEV2.1 tool was obtained. The objective of this work was to evaluate the level of precision and applicability of dCasEV2.1 for achieving plant metabolic engineering interventions. The modulation of the pathways towards the production of a specific metabolite has important implications in plant biomanufacturing since the isolation of pure valuable metabolites could be challenging and laborious. Due to its branched nature and the added value of downstream metabolites, the phenylpropanoid pathway with its derivation into the flavonoid pathway was chosen to be reprogrammed using dCasEV2.1. Initially, individual activations of each gene implicated in the different branches of the flavonoid pathway were performed to optimize the transcriptional activation values obtained with the CRISPRa tool. This was followed by the simultaneous regulation of a set of genes to shift the metabolic flux. The multiplexing capacity of CRISPR/dCas9 previously adapted to this strategy was a determining factor for the success of the approach that enables the creation of complex constructs with a large number of gRNAs targeting different genes in the metabolic pathway. If the dCasEV2.1 tool acts as a signal processor, the gRNA sets that contain the transcriptional regulation instructions can be defined as the activation

---

programs. Four different activation programs for the preferential accumulation of individual flavanones (naringenin or eriodictyol) and flavonols (quercetin or kaempferol) were created. The successful transcriptional outcomes obtained with each activation program were also translated into differences in the metabolic profiles observed in the leaves. A deep non-targeted analysis of the metabolic content of the metabolically reprogrammed leaves revealed a correlation between the activation program employed and the type of flavonoid predominantly accumulated, demonstrating that the dCasEV2.1 offers a high level of precision in multiplexing mode. This approach is suitable to be employed in any metabolic pathway in plants to customize its metabolic content, or for other applications, such as for the understanding of complex genetic networks (Li *et al.*, 2015b). Furthermore, this strategy could be combined with programmable repressors to refine the accuracy of activation programs and avoid unwanted activation of secondary branches. However, currently, there are few examples of efficient repression strategies in plants based on CRISPR. Although not included in this thesis, attempts in our group to add repressor domains to dCasEV2.1 or alternative dCas9 structures have resulted in modest repression rates reaching up to 80% as maximum repression levels. While these levels are encouraging, they are probably not sufficient for many applications in metabolic reprogramming. Some promising results involving alternative nucleases have been reported elsewhere (Tang *et al.*, 2017). Furthermore, classical alternative strategies such as RNAi could be combined with CRISPRa when efficient gene knockdown is required (Small, 2007). Alternatively, emerging epigenetic repressors adding reversible chromatin silencing marks offer new possibilities for targeted gene down-regulation (Gallego-Bartolomé, 2020).

In conclusion, Chapter 2 provides evidence of the high precision achieved with the application of the dCasEV2.1 tool in plants applied in multiplexing mode. Although recent works show the power of other activation strategies based on CRISPR (Xiong *et al.*, 2021; Pan *et al.*, 2021c) that enable efficient simultaneous activation of the genes involved in a complex pathway, the work reported in Chapter 2 has been the first study that has shown the potential of customizing the transcriptional activities in metabolic pathways in plants. Also, the employment of *N. benthamiana* as a chassis for performing metabolic engineering through agroinfiltration has allowed the evaluation of dCasEV2.1



in a fast and versatile way. Transient delivery of genetic information in the form of T-DNA is increasingly popular in plant biotechnology, however other alternative methodologies for genetic delivery have recently emerged to expand the applications of CRISPR strategies in plants (Cody *et al.*, 2017; Ali *et al.*, 2018; Hu *et al.*, 2019).

Finally, Chapter 3 provides an optimized method based on a PVX recombinant virus (Lico *et al.*, 2015), named PVX-VIGR (Virus-Induced Gene Regulation), for the systemic delivery of the gRNAs to the plants, expanding the current toolbox of viral systems employed for transcriptional activation based on TRV (Khakhar *et al.*, 2021). This previously reported strategy provides different configurations to achieve transcriptional repression and activation, showing for the first time an efficient dual transcriptional activity in plants. However, the transcriptional activation and repression levels in this earlier approach were very modest, reaching only between 2 and 5-fold, opening the door to new approaches that provide stronger transcriptional activation rates.

Taking the previously reported work of Uranga *et al.* (2020) as a starting point, where the PVX recombinant virus was employed to deliver gRNAs for gene editing (PVX-VIGE) in *N. benthamiana*, a new version of the viral vector was engineered for PVX-VIGR, now incorporating the gRNA2.1 scaffold required for transcriptional activation. After several optimizations, a spray-based method with a very low optical density requirement was employed efficiently to alter the metabolic profiles of *N. benthamiana* transgenic plants that express the dCasEV2.1 cassette.

The PVX-VIGR experiments targeted three different transcriptional factors (*NbODO1*, *NbMYB1*, and *NbMYB24*) whose likely orthologues in Arabidopsis, petunia, and tomato are involved in metabolic reprogramming, affecting scent production, non-volatile phenylpropanoid compounds, and the activation of the jasmonate pathway (Verdonk *et al.*, 2005; Cheng *et al.*, 2009; Battat *et al.*, 2019). Systemic extension of the transcriptional activation of the targeted TFs was achieved using both local agro-inoculation and spraying approaches. A non-targeted LC-MS analysis revealed a differential metabolic profile for each TF. Although this approach does not allow maintaining the transcriptional activation in the following generations, it allows a rapid evaluation of the metabolic changes produced by the transcriptional perturbation in a

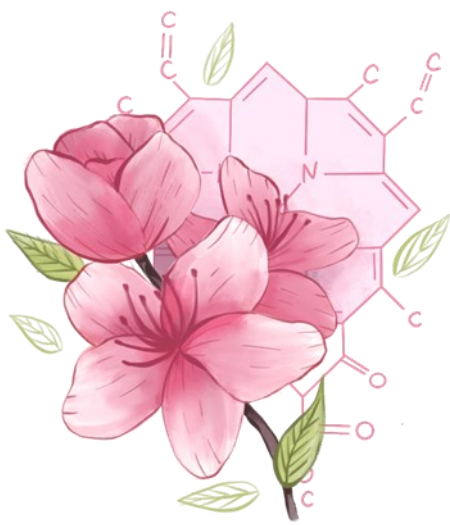
target gene, thus revealing the role of uncharacterized TFs in plants, as was the case of the targeted genes in this study. This is equivalent to the use of VIGS in plant biology, but in this case, the strategy is open to spatiotemporal gene activation as well, opening the door to new functional insights.

In addition to controlling the expression of endogenous genes to change the metabolic profile of the plant, the PVX-VIGA strategy could be employed to control synthetic metabolic pathways whose continued expression is detrimental to plant development. The synthetic promoters used to drive the expression of genes of interest in plants typically employ “natural” TFs that recognize synthetic *cis*-regulatory sequences, however, this may lead to unwanted activities (Petolino and Davies, 2013). The design of synthetic promoters driven by dCasEV2.1 could improve the orthogonality of transgene activation. Some examples of this kind have already been reported in bacteria where synthetic promoters were designed that were activated by dCas9-TFs (Farzadfard *et al.*, 2013). The optimized PVX-VIGR strategy coupled with dCasEV2.1 could offer new approaches to control synthetic metabolic pathways in plants driven by CRISPR/Cas-regulated synthetic promoters.

The novel genetic tools that provide new characteristics and functions to plants of agronomic and biotechnological interest continue to be developed (Park *et al.*, 2015; Bernabé-Orts *et al.*, 2020; Patron, 2020; Torti *et al.*, 2021). The work carried out in this doctoral thesis has addressed the development and optimization of CRISPR tools for transcriptional activation. The dCasEV2.1 strategy has been the result of extensive optimization and evaluation and provides a powerful, specific, and programmable tool for transcriptional activation in plants. The precision of this system has been successfully applied for reprogramming a complex metabolic pathway, followed by the development of a PVX-VIGR strategy for its delivery, which maximizes the possible applications in plants. We expect that the results obtained in this thesis will facilitate the development of new gene circuits that equip plants with favourable synthetic traits, improving their ability to respond to the challenges the world will face.







# Conclusions



## CONCLUSIONS

1. An optimized CRISPR-based programmable transcriptional activator was developed, named dCasEV2.1, that combines the EDLL and VPR protein activation domains with an improved guide RNA scaffold (gRNA2.1) that incorporates two MS2 aptamers in the 3' end of its sequence. The dCasEV2.1 tool was proven to be a strong and easily multiplexable transcriptional activator.
2. Differential expression analysis comparing the transcriptomes of dCasEV2.1-treated and control leaves demonstrates that the dCasEV2.1 tool has very high target specificity when used to activate the endogenous *Nicotiana benthamiana* *NbDFR* gene, with neglectable activation levels of unexpected off-targets in the transcriptome.
3. The multiplexing capacity of the dCasEV2.1 tool enables the activation of two or more genes of choice simultaneously. When the multiplex-activation feature was applied to the induction of genes in the phenylpropanoid pathway, selective reprogramming of the metabolic flux was achieved, leading to a customized activation of different branches of the pathway and resulting in the predominant accumulation of pre-selected compounds, namely naringenin, eriodictyol, kaempferol, and quercetin derivatives.
4. A Viral Induced Gene Activation (VIGA) strategy based on the dCasEV2.1 system and the Potato Virus V (PVX) virus was developed. In this strategy, the constant components of the dCasEV2.1 system, namely the dCas9 enzyme and the MS2:VPR chimaera, were stably integrated into the plant genome, whereas the activation program-containing gRNA element was delivered by agro-infection using a modified PVX viral vector.

5. The use of the PVX-VIGA delivery strategy in *Nicotiana benthamiana* led to strong transcriptional activation of all target genes assayed. The PVX-VIGA-mediated targeted activation of three endogenous *N. benthamiana* MYB genes, whose closest homologues in other species are known to be involved in the regulation of secondary metabolism, resulted in transcriptomic and metabolic changes which were specific for each MYB activation-program assayed, suggesting the ability of PVX-VIGA to re-program the metabolic footprint in adult plants.

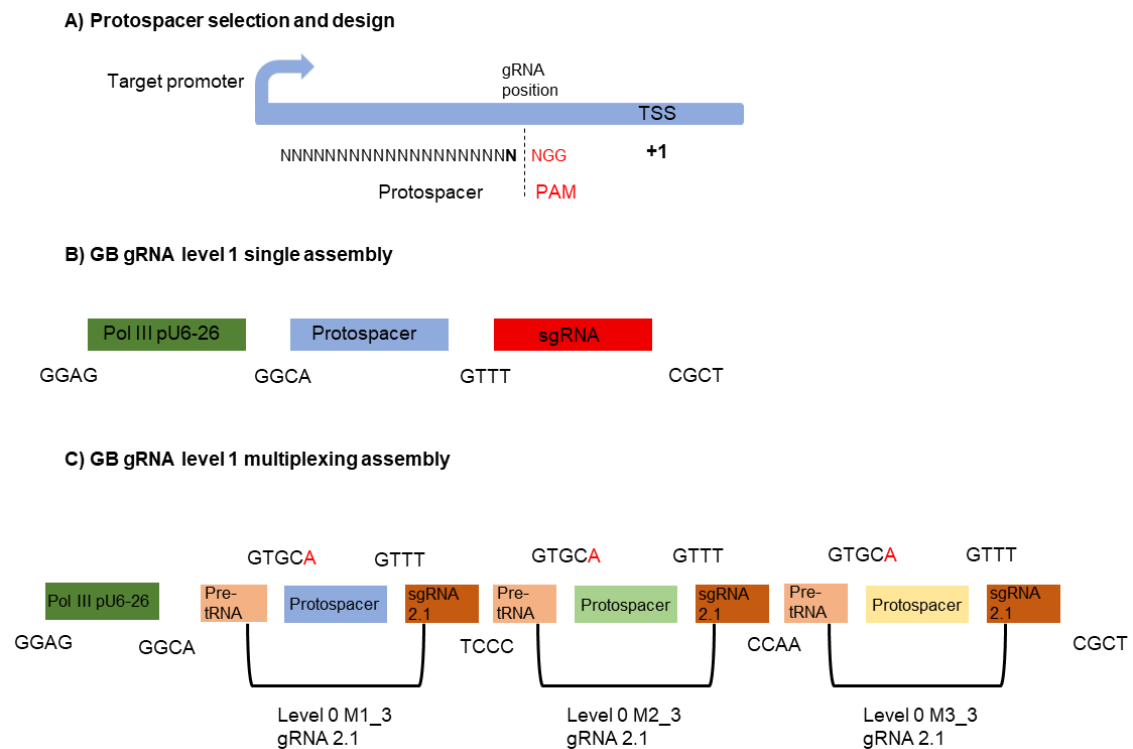


## Supplementary data





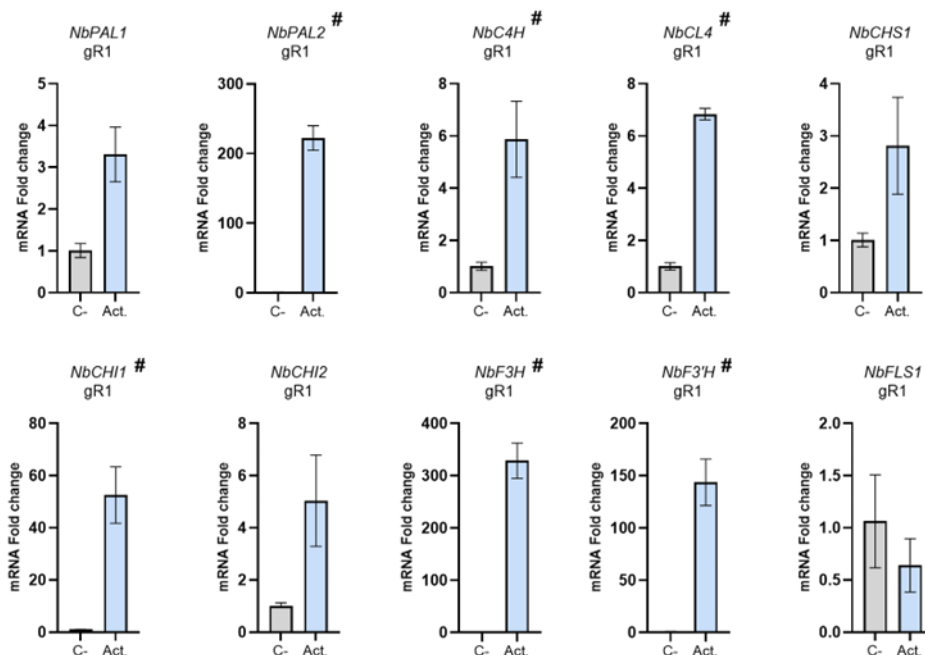
## SUPPLEMENTARY FIGURES



**Supplementary Figure 1: gRNA design and assembly.** A) Protospacer selection for gRNA design. The gRNA position was established taking as a reference the Transcriptional Start Site (TSS) of the target promoter and the first nucleotide after the PAM sequence. B) Representation of the single gRNA GB assembly. C) Representation of the multiplexing strategy for gRNA GB assembly with the gRNA scaffolds 2.1.

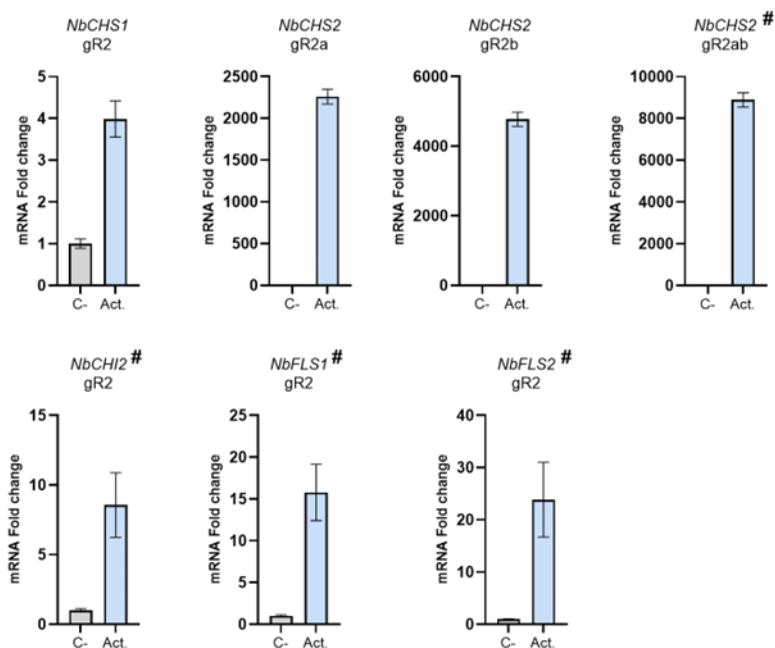
A

**gRNA selection round 1**

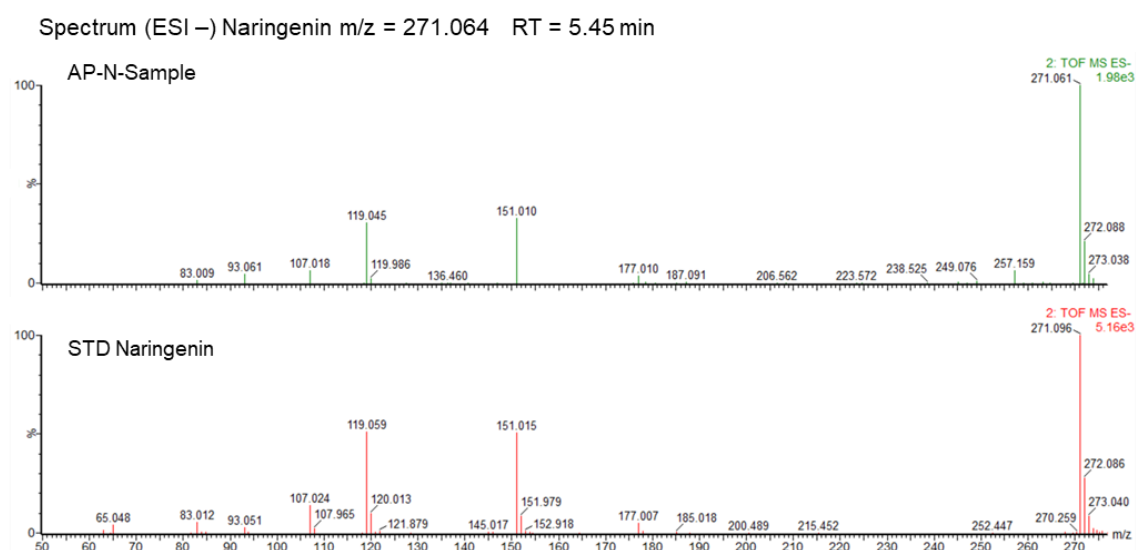


B

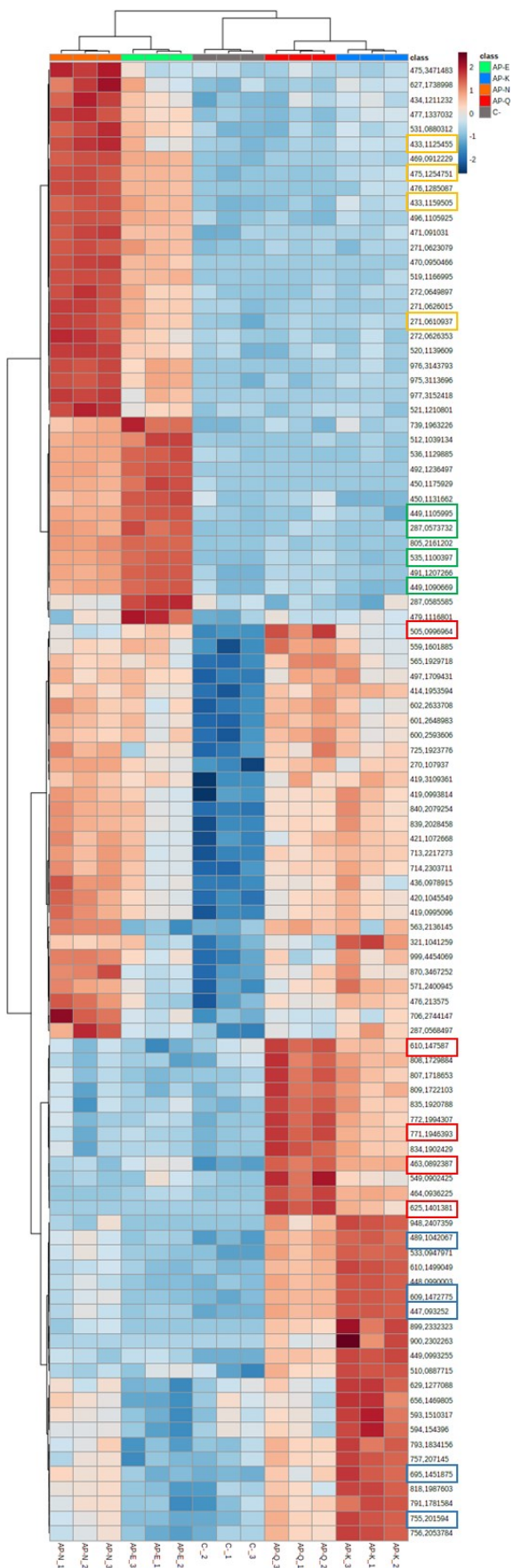
**gRNA selection round 2**



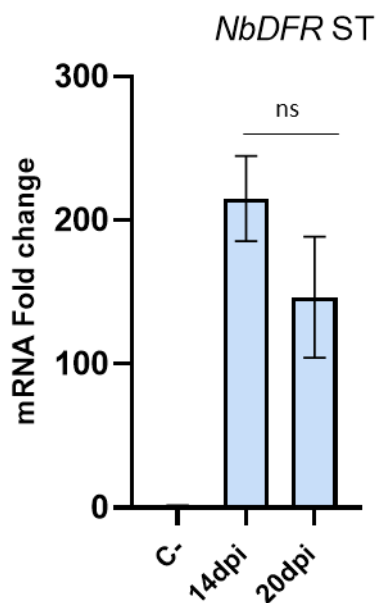
**Supplementary Figure 2: gRNA optimization and selection.** A) mRNA fold change at 4 dpi obtained by targeting the endogenous genes of the flavonoid pathway in the first round of gRNA selection for each gene. B) mRNA fold change at 4 dpi obtained by targeting the endogenous genes of the flavonoid pathway in the second round of gRNA selection for each gene. The # symbol represents the combination of gRNA that were selected for the following assays. Bars represent average fold change  $\pm$  SD, n=3 technical replicates in all experiments.



**Supplementary Figure 2: Negative-ion mode fragmentation of naringenin.** MS-MS spectra of AP-N sample and naringenin commercial standard (STD Naringenin).



**Supplementary Figure 4: Detailed version of hierarchical cluster analysis and heatmap representation of metabolic profile of C- AP-N3, AP-E, AP-K, and G AP-Q samples.** The metabolites represented are the top 100 more significant using a t-test analysis ( $p < 0.05$ ). The data was obtained using Euclidean distance and Ward's minimum variance method. The red boxes indicate up-regulated  $m/z$ . The blue boxes indicate down-regulated  $m/z$ . Each  $m/z$  corresponds to the same compound in all samples. The  $m/z$  selected to quantify the metabolites identified are marked with a square (Orange squares for Naringenin derivatives, green squares for Eriodytiol derivatives, red squares for quercetin derivatives and blue squares for kaempferol derivatives).



**Supplementary Figure 5: Evaluation of PVX-based sgRNA delivery for CRISPRa.** *NbDFR* mRNA fold change at 14 and 20 dpi obtained from the first symptomatic leaf (ST, systemic tissue), by targeting the endogenous gene *NbDFR* with dCasEV2.1 at positions -145 and -198 bp relative to the TSS through the PVX-based sgRNA delivery by syringe inoculation. The C- represents a negative control where a non-specific gRNA was delivered with the PVX-VIGR. Bars represent average RTAs  $\pm$  SD, n=3.



## SUPPLEMENTARY TABLES

Supplementary Table 1. GB level 0 parts used and generated, available in [www.gbcloning.upv.es/](http://www.gbcloning.upv.es/).

Nº GB	Name	GB Nickname
GB1079	pUPD2:dCas9	pdCas9
GB1001	pUPD: U6-26	pU6-26
GB0645	pUPD2: sgRNA	psgRNA
GB1436	pUPD2: sgRNA aptamer SAM	pUPD2 sgRNA:scaffold tetraloop MS2 aptamer
GB2461	pUPD2:sgRNA2.0 scRNA	pUPD2_sgRNA sCRNA 2.0
GB1437	pUPD2: sgRNA2.1 scRNA	pUPD2_sgRNAscaffold 2xF6Ms2 Aptamer
GB1450	pUPD2: sgRNA PP7 aptamer	pUPD2 PP7 Stemloop
GB1451	pUPD2: sgRNA COM aptamer	pUPD2 scf_COM
GB2464	pUPD: SunTag	SunTag
GB1463	pUPD2: ScFv	pUPD2_ScFv
GB1435	pUPD2:MS2	pUPD2_MS2
GB1453	pUPD2:PP7	pUPD2 NLS-PP7
GB1786	pUPD2:COM	pUPD2:NLS-COM1
GB1186	pUPD2:VP64	p3xNLS-VP64
GB1187	pUPD2: EDLL	p3xNLS-EDLL
GB1791	pUPD2: p300 core	pUPD2_P300core
GB1850	pUPD2: VP192	pUPD2_VP192
GB1814	pUPD2: VPR	pUPD2_VPR
GB1817	pUPD2:ERF2(M)	pUPD2_ERF2
GB2001	pUPD2:TV	pUPD2_TV
GB0030	pUPD: p35s	pP35S
GB0037	pUPD:tNos	pTnos
GB2382	pUPD2:NbAN2	pUPD2_NbAN2

**Supplementary Table 2. GB level 1 and >1 TUs and Modules.**

Nº GB	Name	GB database Nickname
<b>GB1398</b>	pNos:Luciferase:tNos-35S:Renilla:tNos-35S:P19:tNos	pEGB3alpha2 Pnos:luc:Tnos-SF-35S:Ren:Tnos-35s:P19:Tnos-SF
<b>GB2248</b>	tNos:NptII:Pnos-Pnos:Luc:tNos-35s:Ren:tNos	Tnos:NptII:Pnos-Pnos:Luc:Tnos-35s:Ren:Tnos
<b>GB1399</b>	SIMTB:luc:tNos-SF-35S:Ren:tNos-35s:P19:tNos-SF	pEGB3alpha2 MTB:luc:Tnos-SF-35S:Ren:Tnos-35s:P19:Tnos-SF
<b>GB2250</b>	tNos:NptII:Pnos-SIDFR:Luc:tNos-35s:Ren:tNos	Tnos:NptII:Pnos-SIDFR:Luc:Tnos-35s:Ren:Tnos
<b>GB1160</b>	SIDFR:Luc:tNos-SF-35S:Renilla:tNos-35S:P19:tNos	pEGB SIDFR:Luc:Tnos-SF-35S:Renilla:Tnos-35S:P19:Tnos
<b>GB2249</b>	tNos:NptII:pNos-SIMTB:Luc:tNos-35s:Ren:tNos	Tnos:NptII:Pnos-MTB:Luc:Tnos-35s:Ren:Tnos
<b>GB0164</b>	p35s:Luciferase:tNos-35S:Renilla:tNos-35S:P19:tNos	35s:Luciferase:Tnos-SF-35s:Renilla:Tnos-35s:P19:Tnos
<b>GB1603</b>	35s:dCas9:SunTag:tNos	35s:dCas9-SunTag:Tnos
<b>GB1189</b>	35s:dCas9:VP64:tNos	pEGB 35s:dCas9-VP64:tNOS
<b>GB1190</b>	35s:dCas9:EDLL:tNos	pEGB 35s:dCas9-EDLL:tNOS
<b>GB1794</b>	35s:dCas9:p300:tNos	3alpha2_35s-dCas9:P300_Tnos
<b>GB1851</b>	35s:dCas9:VP192:tNos	3alpha2_35s-dCas9:VP192-Tnos
<b>GB1826</b>	35s:dCas9:VPR:tNos	3alpha2_35s-dCas9:VPR-Tnos
<b>GB1824</b>	35s:dCas9:ERF2:tNos	3alpha2_35s-dCas9:ERF2-Tnos
<b>GB2047</b>	35s:dCas9:TV:tNos	3alpha2_35s-dCas9:TV-Tnos
<b>GB1403</b>	35S:MS2-VP64:tNos	pDGB3alpha2 35S:MS2-VP64:Tnos
<b>GB1738</b>	35s:MS2:EDLL:tNos	pDGB3_alpha2: 35s_MS2:EDLL_Tnos
Nº GB	Name	GB database Nickanme
<b>GB1398</b>	pNos:Luciferase:tNos-35S:Renilla:tNos-35S:P19:tNos	pEGB3alpha2 Pnos:luc:Tnos-SF-35S:Ren:Tnos-35s:P19:Tnos-SF
<b>GB2248</b>	tNos:NptII:Pnos-Pnos:Luc:tNos-35s:Ren:tNos	Tnos:NptII:Pnos-Pnos:Luc:Tnos-35s:Ren:Tnos
<b>GB1399</b>	SIMTB:luc:tNos-SF-35S:Ren:tNos-35s:P19:tNos-SF	pEGB3alpha2 MTB:luc:Tnos-SF-35S:Ren:Tnos-35s:P19:Tnos-SF

**Supplementary Table 3. Protospacer sequence and position design for each promoter.**

Promoter	Position to TSS	Strand	Sequence
pNos	-161	Coding	GCCACTGAGCCGCGGGTTTC
pNos	-211	Coding	GGGACAAGCCGTTTTACGTT
SIMTB	-98	Coding	TACGATCACGACACGTGTAC
SIMTB	-129	Coding	GATGAAATTAGGATCATGTA
SIMTB	-184	Coding	GTCTAGAACATACGTACGAA
SIMTB	-541	Non coding	GATGTAGCATATGAGATGAT
SIDFR	-150	Non coding	GACTGGTTGGTGAGAGAAGA
SIDFR	-300	Non coding	GGATAAAATGGTAATAGTTT
SIDFR	-376	Coding	GCTGTATCTAATAGAATCTT
NbDFR	-88	Coding	ATGACTGACTGGTTGGTGAG
NbDFR	-125	Non-coding	TCCATATATAGATAAGAAAG
NbDFR	-198	Coding	TATCCGTATGCCTTACCTTT
NbDFR	-217	Coding	TTGGATTTTGGTGTATTCTT
NbDFR	-248	Non coding	TTGAGAATTTGGTAAAACGA
NbAN2	-103	Coding	GGAGTTACGCTAATCACTAG
NbAN2	-145	Non coding	CAGTGCAATTTATTACTCAT
NbAN2	-175	Coding	CGTAAAAAGTCCATATCGAC
NbAN2	-198	Coding	GACGCGTAGCTCTCTCCAAT
NbAN2	-196	Non coding	GAACAGTGTCTACTGCCAAT
NbAN2	-252	Coding	TGTCCTTTTCACTATTAAGT
NbAN1	-101	Non-coding	TAGGAGGAATGAGGTGCGT
NbAN1	-120	Coding	AATAATAGTATAATAACCAA
NbAN1	-173	Coding	TTATACCGTAAGTAACTTAG
NbAN1	-219	Non coding	GTATTACGCTTTTAATCACT
NbAN1	-242	Non coding	CACAACAATCTAATTAATAA
NbAN1	-261	Coding	ACCCGGGTCAAGCCGTGTAA

**Supplementary Table 4. GB gRNA Level 1 for standard and multiplexing strategy.**

<b>Nº GB</b>	<b>Name</b>	<b>GB database nickname</b>
<b>GB1197</b>	U6-26-gRNA-161 pNos	pEGB U626:gRNA4pNOS:sgRNA
<b>GB2462</b>	U6-26-gRNA-161 pNos scf 2.0	3alpha1_U6-26-gRNA4pNos scf aptamer 2.0 Native
<b>GB1724</b>	U6-26-gRNA-161 pNos scf 2.1	pDGB3_alpha1_U6-26:gRNA A 4 (pNos):MS2 F6x2 aptamer
<b>GB1725</b>	U6-26-gRNA-161 pNos scf SAM	pDGB3_alpha1_U6-26:gRNA 4 (pNos):MS2 tetraloop and loop aptamer
<b>GB1740</b>	U6-26-gRNA-161 pNos scf PP7 aptamer	pDGB3_alpha1_U6- 26:gRNA4(pNos):PP7stem loop
<b>GB1797</b>	U6-26-gRNA-161 pNos scf COM aptamer	3alpha1_U6-26-gRNA4pNos COM scf
<b>GB1744</b>	U6-26-gRNA-211 pNos scf 2.0 *	pDGB3_alpha1_U6- 26:gRNA5(pNos):MS2 F6x2 aptamer
<b>GB2049</b>	U6-26-gRNA -161Pnos scf 2.1 - 35s:dCas9:EDLL:Tnos-U6-26-gRNA-211 Pnos scf 2.0* - 35s:MCP:VPR:Tnos	pDGB3_alpha2_U6-26-4gRNA Pnos- F6x2_35s-dCas9:EDLL-Tnos-U6-26- 5gRNA Pnos scf F6x2 - 35s-Ms2:VPR- Tnos
<b>GB2045</b>	U6-26-gRNA-98 SIMTB scf 2.1	3alpha1:U6-26-2gRNA_MTB_scf F6x2
<b>GB1801</b>	U6-26-gRNA-129 SIMTB scf 2.1	3alpha1_U6-26_MTB3-sgRNA-MS2 F6x2 scf
<b>GB2044</b>	U6-26-gRNA-184 SIMTB scf 2.1	3alpha2:U6-26-4gRNA_MTB_scf F6x2
<b>GB1859</b>	U6-26-gRNA-541 SIMTB scf 2.1	3alpha1:U6-26-5gRNA_MTB_scf F6x2
<b>GB2070</b>	U6-26-gRNA-541 SIMTB scf 2.1 + U6-26- gRNA-541 SIMTB scf 2.1 + U6-26-gRNA- 129 SIMTB scf 2.1 + U6-26-gRNA-98 SIMTB scf 2.1	pDGB3_Alpha1_U6-26-5gRNA MTB- F6x2_U6-26-4gRNA MTB-F6x2_U6-26- 3gRNA MTB-F6x2_U6-26-2gRNA MTB- F6x2
<b>GB1838</b>	U6-26-gRNA-150 SIDFR scf 2.1	3alpha1_U6-26-1gRNA-DFR F6x2 MS2scf
<b>GB1837</b>	U6-26-gRNA-300 SIDFR scf 2.1	3alpha1_U6-26-4gRNA-DFR F6x2 MS2scf

Table S4 (continued).

<b>N° GB</b>	<b>Name</b>	<b>GB database Nickname</b>
<b>GB1839</b>	U6-26-gRNA-376 SIDFR scf 2.1	3alpha1_U6-26-5gRNA-DFR F6x2 MS2scf
<b>GB2169</b>	Multiplex scRNA 2.1 U6-26 gRNA -88, -125,-217 NbDFR	3Alpha1_U6-26-sgRNANbDFR -85, -145,-218 F6x2 Multiplex
<b>GB2170</b>	Multiplex scRNA 2.1 U6-26 gRNA -88, -198,-248 NbDFR	3Alpha2_U6-26-sgRNANbDFR -85, -198,-268 F6x2 Multiplex
<b>GB2176</b>	Multiplex scRNA 2.1 U6-26 gRNA -88, -125,-217 NbDFR + Multiplex scRNA 2.0 U6-26 gRNA -88, -198,-248 NbDFR	3Omega2_U6-26-sgRNANbDFR -85, -145,-218 + U6-26-sgRNANbDFR -85, -198,-268 F6x2 Multiplex
<b>GB2171</b>	Multiplex scRNA 2.1 U6-26 gRNA -103, -175,-196 NbAN2	3Alpha1_U6-26-sgRNANbAN2 -103, -175,-196 F6x2 Multiplex
<b>GB2172</b>	Multiplex scRNA 2.1 U6-26 gRNA -125, -198,-252 NbAN2	3Alpha2_U6-26-sgRNANbAN2 -125, -198,-252 F6x2
<b>GB2177</b>	Multiplex scRNA 2.1 U6-26 gRNA -103, -175,-196 NbAN2 + Multiplex scRNA 2.0 U6-26 gRNA -125, -198,-252 NbAN2	3Omega2_U6-26-sgRNANbAN2 -103, -175,-196 + U6-26-sgRNANbAN2 -125, -198,-252 F6x2
<b>GB2309</b>	Multiplex scRNA 2.1 U6-26 gRNA -101, -173,-242 NbAN1	3alpha 1 Multiplex U626_NbAN14283-101-173-242
<b>GB2310</b>	Multiplex scRNA 2.1 U6-26 gRNA -120, -219,-261 NbAN1	3alpha 2 Multiplex U626_NbAN14283-120-219-261
<b>GB2408</b>	Multiplex scRNA 2.1 U6-26 gRNA -101, -173,-242 NbAN1 + Multiplex scRNA 2.0 U6-26 gRNA -120, -219,-261 NbAN1	Omega2:U2626_NbAN1_4283_-101-173-242 F6 + U2626_NbAN1_4283_-120-219-261 F6

**Supplementary Table 5. GB level -1 parts for multiplexing strategy.**

Nº GB	Name	GB database nickname
<b>GB2073</b>	pVD1_M1-3pTRNA scf 2.1	pVD1_M1-F6-pTRNA-scf
<b>GB2074</b>	pVD1_M2-3pTRNA scf 2.1	pVD1_M2-F6-pTRNA-scf
<b>GB2075</b>	pVD1_M3-3pTRNA scf 2.1	pVD1_M3-F6-pTRNA-scf

**Supplementary Table 6. GB gRNAs level 0 for multiplexing assembly.**

Nº GB	Name	GB database nickname
<b>GB2148</b>	Multiplex M1_3gRNA-88 NbDFR 2.1	pUPD2_M1-F6-sgRNANbDFR-85
<b>GB2149</b>	Multiplex M2_3gRNA-125 NbDFR 2.1	pUPD2_M2-F6-sgRNANbDFR-145
<b>GB2150</b>	Multiplex M2_3gRNA-198 NbDFR 2.1	pUPD2_M2-F6-sgRNANbDFR-198
<b>GB2151</b>	Multiplex M3_3gRNA-217 NbDFR 2.1	pUPD2_M3-F6-sgRNANbDFR-218
<b>GB2152</b>	Multiplex M3_3gRNA-248 NbDFR 2.1	pUPD2_M3-F6-sgRNANbDFR-268
<b>GB2153</b>	Multiplex M1_3gRNA-103 NbAN2 2.1	pUPD2_M1-F6-sgRNANbAN2-103
<b>GB2154</b>	Multiplex M1_3gRNA-125 NbAN2 2.1	pUPD2_M1-F6-sgRNANbAN2-125
<b>GB2155</b>	Multiplex M2_3gRNA-175 NbAN2 2.1	pUPD2_M2-F6-sgRNANbAN2-175
<b>GB2156</b>	Multiplex M2_3gRNA-198 NbAN2 2.1	pUPD2_M2-F6-sgRNANbAN2-198
<b>GB2157</b>	Multiplex M3_3gRNA-196 NbAN2 2.1	pUPD2_M3-F6-sgRNANbAN2-196
<b>GB2158</b>	Multiplex M3_3gRNA-252 NbAN2 2.1	pUPD2_M3-F6-sgRNANbAN2-252
<b>GB2302</b>	Multiplex M1_3gRNA-101 NbAN1 2.1	pUPD2_M1-3NbAN1_4283-101
<b>GB2303</b>	Multiplex M1_3gRNA-120 NbAN1 2.1	pUPD2_M1-3NbAN1_4283-120
<b>GB2304</b>	Multiplex M2_3gRNA-173 NbAN1 2.1	pUPD2_M2-3NbAN1_4283-173
<b>GB2305</b>	Multiplex M2_3gRNA-219NbAN1 2.1	pUPD2_M2-3NbAN1_4283-219
<b>GB2306</b>	Multiplex M3_3gRNA-242 NbAN1 2.1	pUPD2_M3-3NbAN1_4283-242
<b>GB2307</b>	Multiplex M3_3gRNA-261 NbAN1 2.1	pUPD2_M1-3NbAN1_4283-261

**Supplementary Table 7. Primer pair for qRT analysis.**

Target Gene	Sequence 5' - 3'
NbDFR	TTCATCTGCGCATCCCATCA
	TCCCTACTGAGTTTAAAGGTATCGA
NbAN1	CATCTCTTAATAATGGCGTCTTCTTG
	CTCTAGGGATTATCTGATGTATTGACC
NbAN2	GGAAAAGTTGCAGACTGAGGTG
	ACCCGCAATAAGTGACCATCTG
NbF-box	TTGGAAACTCTCTCCCCACTTG
	GCTCATTGTTGGATGGGTACCT

**Supplementary Table 8. Potential off-targets of gRNAs used for *NbDFR* dCas9EV2.1 activation.**

gRNA position	Targeted gene	Potential off-targets	Off-target found
NbDFR -88	Niben101Scf00305g05035	Niben101Scf00311g05015(UP) Niben101Scf00606g02015(UP) Niben101Scf05634g02005(UP)	Niben101Scf00606g02015
NbDFR -125	Niben101Scf00305g05035	Niben101Scf01792g00022(UP) Niben101Scf02399g01012(DOWN) Niben101Scf02764g05013(UP) Niben101Scf02764g05013(DOWN) Niben101Scf03455g04002(DOWN) Niben101Scf05217g01001(DOWN) Niben101Scf14642g04009(DOWN)	NONE
NbDFR -198	Niben101Scf00305g05035	Niben101Scf01607g06019(UP) Niben101Scf21739g00001(UP)	NONE
NbDFR -218	Niben101Scf00305g05035	Niben101Scf00320g02010(UP) Niben101Scf00320g02006(UP)	
NbDFR -248	Niben101Scf00305g05035	Niben101Scf00606g02015(UP)	Niben101Scf00606g02015

**Supplementary Table 9: Potential off-targets of gRNAs used for *NbAN2* dCas9EV2.1 activation.**

<b>gRNA position</b>	<b>Targeted gene</b>	<b>Potential off-targets</b>	<b>Off-target found</b>
NbAN2 -103	Niben101Scf00156g02004	Niben101Scf00285g10004(DOWN)	Niben101Scf00285g10004
NbAN2 -145	Niben101Scf00156g02004	Niben101Scf00152g10003(UP) Niben101Scf00156g03002(UP) Niben101Scf00285g10004(DOWN) Niben101Scf02868g03005(UP) Niben101Scf03306g00002(DOWN) Niben101Scf08512g00009(DOWN)	Niben101Scf00285g10004
NbAN2 -175	Niben101Scf00156g02004	Niben101Scf00285g10004(DOWN) Niben101Scf00288g15003(DOWN) Niben101Scf03414g03001(UP) Niben101Scf03518g00006(UP)	Niben101Scf00285g10004
NbAN2 -198	Niben101Scf00156g02004	Niben101Scf00285g10004(DOWN)	Niben101Scf00285g10004
NbAN2 -196	Niben101Scf00156g02004	Niben101Scf00285g10004(DOWN)	Niben101Scf00285g10004
NbAN2 -252	Niben101Scf00156g02004	Niben101Scf01383g09023(UP) Niben101Scf05342g05013(UP) Niben101Scf14009g01001(UP)	NONE



**Supplementary Table 10. List of candidate flavonoid genes identified in this work in *N. benthamiana*.** Gene accessions for solgenomics.com and nbentham.com are provided. The protospacer sequences designed for each round of optimization are included for all tested genes.

Gene	Solgenomics Accession	Nbentham.com Accession	gRNAs sequence gR1	gRNAs sequence gR2
<i>NbPAL1</i>	Niben101Scf04 656g01001	NbLab330C12: NbE44070663.1	gRNA1: CCTAATTAATATCTTGAAA gRNA2: CAATAATGTTTTTCAGAAAA gRNA3: TCACCACTATAATATAAGA	
<i>NbPAL2</i>	Niben101Scf02 432g00011	NbLab330C16: 80768655...807 71248	gRNA1: TGAAGAAGCTAGACTGACCGT gRNA2: GTTGAAGGGTGGTTGTTGAG gRNA3: ATATTGTTTCATATTATTAC	
<i>NbCHS1</i>	Niben101Scf02 893g00001.1	NbLab330C10: NbD021251.1	gRNA1: TAAITCTTAAGATTTCTAAT gRNA2: TCCTAAAAATTATAATGTATG gRNA3: GGATTTGATTTTAAAGATAT	gRNA1: TAAITCTTAAGATTTCTAAT gRNA4: GAAGTGCATTATGTTTGCAT gRNA5: AGAAAAATATCATTTTGCAAGA
<i>NbCHS2</i>	Niben101scf00 536g13015	NbLab330C10: NbD004910.1		gR2a: gRNA1: AGGTCACGTGATCTTAATGT gRNA2: TGCTGAGAGAAGAGTTGAAA gRNA3: TACCTACATCTCAAGATCAA gR2b: gRNA1: AGGTCACGTGATCTTAATGT gRNA4: AATCTATTGAATCAAATATA gRNA5: CCAAAGCATCAITCTTCAAA
<i>NbCHI1</i>	Niben101Scf05 989g01008	NbLab330C11: NbD035233.1	gRNA1: AAAGTTGGGTGAATCCTTTA gRNA2: GTCTATATAAATGTTACTTA gRNA3: ATGAAGGTAACCGATAGAAT	
<i>NbCHI2</i>	NbLab330C10: NbD025871.1	NbLab330C01: NbD014654.1	gRNA1: ACAAGCTGTTAAGGAAAGTT gRNA2: TTAAGAAAAAGTTACTCGTG gRNA3: TATCTTTTCAAAAAAATCT	gRNA1: ACAAGCTGTTAAGGAAAGTT gRNA2: TTAAGAAAAAGTTACTCGTG gRNA4: CGAGAGAGAGGGTTGACTGG
<i>NbC4H</i>	Niben101Scf08 196g01007	NbLab330C00: NbD041626.1	gRNA1: AACCTAAGCAGAAAAATGACG gRNA2: ATAACCTTGTAAATGAAA gRNA3: TTAAGGATAAAAAAGCTTAT	

Table S10 (continued).

Gene	Solgenomics Accession	Nbentn.com Accession	gRNAs sequence gR1	gRNAs sequence gR2
<i>Nb4CL</i>	Niben101Scf07 623g01036	NbLab330C07: NbD040166.1	gRNA1: TAGACATATTAGAAGACAGG gRNA2: TATTTCTATCTTGGGTAAAG gRNA3: TGAGAGTAGTTCAGGGGAATG	
<i>NbF3H</i>	Niben101Scf09 345g00004	NbLab330C02: NbD044172.1	gRNA1: TCCATTGCTTCTGAGTTGAA gRNA2: TATTTTGTAAATTAATAATTT gRNA3: ACGGAGGAATACACTATTTT	
<i>NbF3' H</i>	Niben101Scf00 974g01018	NbLab330C06: NbD007916.1	gRNA1: TTTTTTGGAAATCGACTAGAA gRNA2: AAGAAGTAGATTTTCAAAAT gRNA3: AAATATTTAAAGCACGAGGT	
<i>NbFLS1</i>	Niben101Scf02 429g06007	NbLab330C03 NbD018272.1	gRNA1: ATTTTGAATAAAATATGTA gRNA2: TTAATTTTATCTTGATGAA gRNA3: ATCATATTTCTAAAAGAACG	gRNA4: GCCTATACCTAACACTCTCA gRNA5: TTTCTCGTGGGTGACTTACG gRNA6: TTACACGAAACAAGTGAAGCT
<i>NbFLS2</i>	Niben101Scf03 779g04001	NbLab330C10 NbD025871.1		gRNA1: GTACATTCATGTGAACATGT gRNA2: GGAATGACTAGGATTGAAAA gRNA3: TCTTCATATTCGTTTTCTGT
<i>NbDFR</i>	Niben101Scf00 305g05035	NbLab330C19: NbD003043.1	gRNA1: ATGACTGACTGGTTGGTGAG gRNA2: TATCCGTATGCCTTACCTTT gRNA3: TTGAGAAATTTGGTAAAAACGA	

**Supplementary Table 11: List of GoldenBraid plasmids used in this work.** DNA sequences can be found at <https://gbcloning.upv.es/> by entering the provided GB\_ID.

<b>GB_ID</b>	<b>Construct</b>	<b>GB</b>	<b>Construct</b>
GB2085	dCasEV	GB2394	NbFLS1 gR1
GB2070	Non-target gRNA2.1	GB2639	NbFLS1 gR2
GB2389	NbPAL1 gR1	GB2530	NbFLS2 gR2
GB2396	NbPAL2 gR1	GB2170	NbDFR gR1
GB2760	NbCL4 gR1	GB2641	AP-N3
GB2390	NbCHS1 gR1	GB2777	AP-N1
GB2500	NbCHS2 gR2a	GB2776	AP-N1B
GB2397	NbCHS2 gR2b	GB2864	AP-N2
GB2599	NbCHS2 gR2ab	GB2863	AP-N2B
GB2502	NbCHI1 gR1	GB2866	AP-N4
GB2391	NbCHI2 gR1	GB2867	AP-N4B
GB2503	NbCHI2 gR2	GB3259	AP-K
GB2392	NbF3H gR1	GB3242	AP-E
GB2531	NbF3'H gR1	GB3243	AP-Q
GB2395	NbC4H gR1		

**Supplementary Table 12: Primer pairs for RT-qPCR analyses.**

Gene	Primer Sequence5'-3'
<i>NbPAL1</i>	Fw: CGACAAGAGCGGCAATGCTAGTGAG RV: CGGAGAGGCAAGCATGGGGTGATAT
<i>NbPAL2</i>	Fw: AAGATCGCAAAGCACATTTATTTG RV: CGAAGTATATGTGATGTTTTGATGGA
<i>NbCHS1</i>	Fw: GCATTCCAACCATTAGGTCTTTCGGA RV: CGAAGTTTCTCTGGCTTAGGCTCAA
<i>NbCHS2</i>	Fw: ACTACTGGTGAAGGGCTTGAATGGG RV: GTCCTGCCCACTAAGTAGCAACT
<i>NbCHI1</i>	Fw: AATCCTATGAGACAACCTGTTGGCCC RV: TGCATTTACATGCTTGAGTTGACC
<i>NbCHI2</i>	Fw: CCGGATAGGAATTGGCTAAGATCAT RV: TCTTTTCTCCTCGAGAGTTAAGGTC
<i>NbC4H</i>	Fw: GTGTGGGACTAAAAGAGGGATTGCC RV: GAGTAGGAGTGCAAATCACTGAGCC
<i>Nb4CL</i>	Fw: GGGCCATTTGTGCCGCTATTGTTT RV: GCAGCCCCAGACATGACAGTCCTTAC
<i>NbF3H</i>	Fw: GGATTACTGTTCAGCCCGTTGAAGG RV: CTGCTGCTATTTCGAGTTCACCACTG
<i>NbF3'H</i>	Fw: AGAGGGGTGGATGGATGAAGCCTTA RV: CAAAGTGGTTGAAAGAACCGGCTC
<i>NbDFR</i>	Fw: TTCATCTGCGCATCCCATCA RV: TCCCTACTGAGTTTAAAGGTATCGA
<i>NbFLS1</i>	Fw: GCAGAAGGAGGCCAAATCTTATGGAAC RV: GAGATGATTGCAAAGAATCCAGGGTCG
<i>NbFLS2</i>	Fw: TTAGGCCTTCCCCACATTCGGATC RV: CTGGCTTTACAGTGATCCAAACGCC

**Supplementary Table 13. Metabolite tentative identification in AP-N, AP-E, AP-K, and AP-Q samples.** The *m/z* represented were identified as significantly different in the hierarchical clustering and heatmap. The *m/z* marked in bold represents the parental ion in negative ESI used for quantification. The *m/z* were grouped by retention time and checked with the second fragmentation (F2) of the chromatogram for the correct identification. (STD)\* represents the metabolites identified with purified standards.

Most abundant in AP:	<i>m/z</i> detected in F1	Retention Time (min)	Fragmentation of the parental ion (F2)	Tentative identification
AP-N3	469,091	4,31	469, 433, 271, 193, 151	Naringenin Hexose I
AP-N3	271,062			
AP-N3	471,091			
AP-N3	531,088			
AP-N3	627,173			
AP-N3	<b>433,115</b>			
AP-N3	496,11			
AP-N3	<b>433,112</b>	4,49	433, 271, 151, 109	Naringenin Hexose II
AP-N3	271,062	5,15	475, 271, 151	Naringenin Glucoside I
AP-N3	475,347			
AP-N3	476,128			
AP-N3	976,314			
AP-N3	520,113			
AP-N3	977,315			
AP-N3	519,116			
AP-N3	477,133			
AP-N3	521,121			
AP-N3	975,311			
AP-N3	<b>475,125</b>			
AP-N3	<b>271,061</b>	5,37	151, 119	Naringenin (STD)*
AP-E	287,058	3,9	739, 512, 547, 485 449, 287, 151	Eriodictyol Glucoside I
AP-E	<b>449,109</b>			
AP-E	450,113			
AP-E	512,103			
AP-E	805,216			

Table S13 (continued).

Most abundant in AP:	m/z detected in F1	Retention Time (min)	Fragmentation of the parental ion (F2)	Tentative identification
AP-E	450,113	4,58	287	Eriodictyol Hexose
AP-E	<b>449,11</b>			
AP-E	536,114	4,8	535, 491, 287, 151	Eriyodictyol Glucoside II
AP-E	<b>535,109</b>			
AP-E	492,123			
AP-E	491,121			
AP-E	<b>287,057</b>	4,88	151	Eriodictyol (STD)*
AP-E	301,061	5,15	287, 271, 151	Homoeriodictyol / Hespertin
AP-K	900,23	3,67	609, 579 447, 285	Kaempferol Dihexose
AP-K	<b>609,147</b>			
AP-K	899,233			
AP-K	610,149			
AP-K	948,24			
AP-K	818,198	3,75	593, 449, 289, 285	Kaempferol Dihexose-dioxyhexose
AP-K	793,183			
AP-K	<b>755,201</b>			
AP-K	756,205			
AP-K	757,207			
AP-K	791,178			
AP-K	<b>695,145</b>	4,3	465, 289, 285, 245	Kaempferol Glucoside II
AP-K	<b>447,093</b>	4,93	447, 285, 133, 115	Kaempferol Hexose
AP-K	448,099			
AP-K	449,099			
AP-K	629,127			
AP-K	593,151			
AP-K	594,154			
AP-K	656,146			
AP-K	510,088			
AP-K	533,094	5,7	489, 433, 285, 173	Kaempferol Glucoside I
AP-K	<b>489,104</b>			
AP-Q	<b>625,14</b>	3,38	609, 465, 301, 299	Quercetin-dihexose

Table S13 (continued).

Most abundant in AP:	m/z detected in F1	Retention Time (min)	Fragmentation of the parental ion (F2)	Tentative identification
AP-Q	835,192	3,45	609, 301	Quercetin Dihexose-dioxyhexose
AP-Q	834,19			
AP-Q	807,171			
AP-Q	<b>771,194</b>			
AP-Q	772,199			
AP-Q	808,172			
AP-Q	809,172			
AP-Q	<b>609,145</b>	4,55	463, 301, 300	Quercetin Hexose-dioxyhexose
AP-Q	610,147			
AP-Q	<b>463,089</b>	4,58	301, 3	QuercetinHexose
AP-Q	464,093			
AP-Q	549,09	5,34	537, 505, 469, 301, 300	Quercetin Glucoside I

Supplementary Table 14. Protospacer sequences designed for targeting the *N. benthamiana* genes with dCasEV2.1.

Gene	Protospacer sequence
<i>NbDFR</i>	gRNA1: ATGACTGACTGGTTGGTGAG gRNA2: TATCCGTATGCCTTACCTTT
<i>NbODO1</i>	gRNA1: TTGGGTGGAGCTGATAACAC gRNA2: TGAATAAGAAAAAAGGAAAG
<i>NbMYB21</i>	gRNA1: ATTTTATAGTAAAGCAAGTT gRNA2: CTATTGATGTTATCATGGCA
<i>NbMYB24</i>	gRNA1: AATTCAGTCTTTAATCAGCC gRNA2: ATTAGCAGTATATCGACTTA

**Supplementary Table 15. Primers employed for sgRNA2.1 construction and adaptation for PVX recombinant plasmids through Gibson assembly.**

Construct	Primer sequence 5' → 3'
PVX::gDFR	Fw:gaggtcagcaccagctagcaATGACTGACTGGTTGGTGAGGTTTTAGAGCTAGAAATAGC Rv: catacggataGGGAAGACTCCCCAGTGAC  Fw: gagtcttcccTATCCGTATGCCTTACCTTTGTTTTAG Rv: gggaaacttaacaaaccctaGGGAAGACTCCCCAGTGAC
PVX::gODO1	Fw: gaggtcagcaccagctagcaAGCATCAACAGGTCTTAAGCGTTTTAG Rv: tagtgtaaataGGGAAGACTCCCCAGTGA  Fw: gagtcttcccATTTACTAACAATAGTAGGTTTTAGA Rv: gggaaacttaacaaaccctaGGGAAGACTCCCCAGTGAC
PVX::gMYB21	Fw: gaggtcagcaccagctagcaATTTTATAGTAAAGCAAGTTGTTTTAG Rv: acatcaatagGGGAAGACTCCCCAGTGAC  Fw: gagtcttcccCTATTGATGTTATCATGGCAGTTTTAGAG Rv: gggaaacttaacaaaccctaGGGAAGACTCCCCAGTGAC
PVX::gMYB24	Fw: gaggtcagcaccagctagcaAATTCAGTCTTTAATCAGCCGTTTTAG Rv: atactgctaataGGGAAGACTCCCCAGTGAC  Fw: gagtcttcccATTAGCAGTATATCGACTTAGTTTTAGAG Rv: gggaaacttaacaaaccctaGGGAAGACTCCCCAGTGAC
PVX::gDFR	Fw:gaggtcagcaccagctagcaATGACTGACTGGTTGGTGAGGTTTTAGAGCTAGAAATAGC Rv: catacggataGGGAAGACTCCCCAGTGAC  Fw: gagtcttcccTATCCGTATGCCTTACCTTTGTTTTAG Rv: gggaaacttaacaaaccctaGGGAAGACTCCCCAGTGAC



**Supplementary Table 16. Primer pairs for RT-qPCR analyses.**

Gene	Primer sequence 5' → 3'
<i>NbDFR</i>	Fw: TTCATCTGCGCATCCCATCA Rv: TCCCTACTGAGTTTAAAGGTATCGA
<i>NbODO1</i>	Fw: AATTA AAAACTGTCGTCCCAAG Rv: GTCTGTCTCAGACCTTCTGCTG
<i>NbMYB21</i>	Fw: TGCCAGGAAGAACAGATAACGAGA Rv: ACGACATATGGCTACTGCTTCCT
<i>NbMYB24</i>	Fw: ATCATCAACAAGCTAGTACAAGC Rv: TCATTTATTTTCGTAAACAATTCATGGT
<i>NbF-box</i>	Fw: GGCACTCACAAACGTCTATTTTC Rv: ACCTGGGAGGCATCTGCTTAT

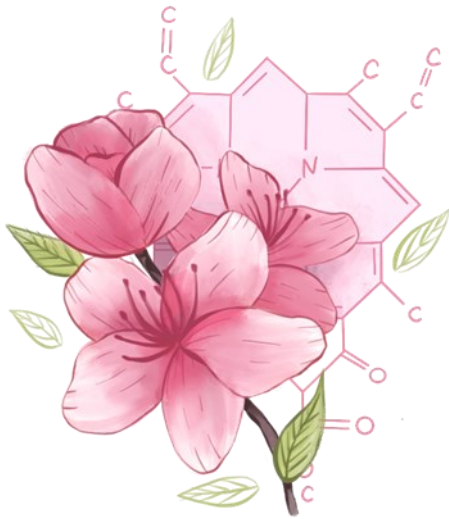
**Supplementary Table 17. m/z fragmentation spectrum F2 of the differential metabolites obtained in the non-targeted analysis of the volatile content in the negative control, *NbODO1*, *NbMYB21*, and *NbMYB24*-induced samples.**

Compound name	Fragmentation F2
2,4-Diisopropylphenyl acetate	178, 163, 117, 147, 105, 91
Methoxycalamene	217, 200, 157, 141
Phenol 3,5 dimethyl	122, 121, 107, 79
3,4,4-Trimethyl-2-cyclopenten-1-one	109, 79, 81, 55, 70, 41
Naphthalene, 1,2-dihydro-1,1,6-trimethyl	172, 163, 157, 142, 115, 128
Naphthalene, 1,2-dihydro-4,5,7-trimethyl	173, 157, 142
$\alpha$ -Terpineol	136, 121, 93, 81, 59
$\alpha$ -Terpinene	136, 121, 93, 91, 77
o-Cymene	207, 119, 134, 91, 57
$\alpha$ ,4-Dimethyl-3-cyclohexene-1-acetaldehyde	177, 135, 91, 43

**Supplementary Table 18. Ct values obtained from the RT-qPCR assay for evaluating the transcriptional expression of *NbODO1*, *NbMYB21*, and *NbMYB24* pPVX-gRNA induced samples.**

<b>Assay</b>	<b>Reference Gene (F-Box)</b>	<b>C- sample (TF measurement)</b>	<b>TF-induced sample</b>
NbODO1	25,36 ± 0,58	31,23 ± 0,41	21,48 ± 1,79
NbMYB21	25,94 ± 0,12	32,91 ± 0,70	25,40 ± 0,60
NbMYB24	25,06 ± 0,23	34,88 ± 0,92	24,43 ± 0,22

# References





## REFERENCES

- Ali Z, Abul-faraj A, Li L, et al.** 2015. Efficient Virus-Mediated Genome Editing in Plants Using the CRISPR/Cas9 System. *Molecular plant* **8**, 1288–1291.
- Ali Z, Eid A, Ali S, Mahfouz MM.** 2018. Pea early-browning virus-mediated genome editing via the CRISPR/Cas9 system in *Nicotiana benthamiana* and *Arabidopsis*. *Virus Research* **244**, 333–337.
- Amack SC, Antunes MS.** 2020. CaMV35S promoter – A plant biology and biotechnology workhorse in the era of synthetic biology. *Current Plant Biology* **24**, 100179.
- An F, Zhao Q, Ji Y, et al.** 2010. Ethylene-Induced Stabilization of ETHYLENE INSENSITIVE3 and EIN3-LIKE1 Is Mediated by Proteasomal Degradation of EIN3 Binding F-Box 1 and 2 That Requires EIN2 in *Arabidopsis*. *The Plant Cell* **22**, 2384–2401.
- Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA, Arkin AP, Keasling JD.** 2010. BglBricks: A flexible standard for biological part assembly. *Journal of Biological Engineering* **4**, 1.
- Antunes MS, Ha S-B, Tewari-Singh N, Morey KJ, Trofka AM, Kugrens P, Deyholos M, Medford JI.** 2006. A synthetic de-greening gene circuit provides a reporting system that is remotely detectable and has a re-set capacity. *Plant biotechnology journal* **4**, 605–622.
- Antunes MS, Morey KJ, Smith JJ, et al.** 2011. Programmable Ligand Detection System in Plants through a Synthetic Signal Transduction Pathway. *PLOS ONE* **6**, e16292.

- Aoyama T, Chua N-H.** 1997. A glucocorticoid-mediated transcriptional induction system in transgenic plants. *The Plant Journal* **11**, 605–612.
- Ariga H, Toki S, Ishibashi K.** 2020. Potato Virus X Vector-Mediated DNA-Free Genome Editing in Plants. *Plant and Cell Physiology* **61**, 1946–1953.
- Arya SS, Mahto BK, Ramkumar TR, Lenka SK.** 2020. Sharpening gene editing toolbox in Arabidopsis for plants. *Journal of Plant Biochemistry and Biotechnology* **29**, 769–784.
- Bae S, Park J, Kim J-S.** 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics (Oxford, England)* **30**, 1473–1475.
- Bai J, Wang X, Wu H, Ling F, Zhao Y, Lin Y, Wang R.** 2020. Comprehensive construction strategy of bidirectional green tissue-specific synthetic promoters. *Plant Biotechnology Journal* **18**, 668–678.
- Balboa D, Weltner J, Eurola S, Trokovic R, Wartiovaara K, Otonkoski T.** 2015. Conditionally Stabilized dCas9 Activator for Controlling Gene Expression in Human Cell Reprogramming and Differentiation. *Stem cell reports* **5**, 448–459.
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**, 1709–1712.
- Battat M, Eitan A, Rogachev I, Hanhineva K, Fernie A, Tohge T, Beekwilder J, Aharoni A.** 2019. A MYB Triad Controls Primary and Phenylpropanoid Metabolites for Pollen Coat Patterning. *Plant physiology* **180**, 87–108.

- Beerli RR, Segal DJ, Dreier B, Barbas CF 3rd.** 1998. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14628–14633.
- Belcher MS, Vuu KM, Zhou A, Mansoori N, Agosto Ramos A, Thompson MG, Scheller HV, Loqué D, Shih PM.** 2020. Design of orthogonal regulatory systems for modulating gene expression in plants. *Nature Chemical Biology* **16**, 857–865.
- Bell-Pedersen D, Cassone VM, Earnest DJ, Golden SS, Hardin PE, Thomas TL, Zoran MJ.** 2005. Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nature Reviews Genetics* **6**, 544–556.
- Benner SA, Sismour AM.** 2005. Synthetic biology. *Nature Reviews Genetics* **6**, 533–543.
- Bernabé-Orts JM, Quijano-Rubio A, Vazquez-Vilar M, Mancheño-Bonillo J, Moles-Casas V, Selma S, Gianoglio S, Granell A, Orzaez D.** 2020. A memory switch for plant synthetic biology based on the phage  $\phi$ C31 integration system. *Nucleic Acids Research* **48**, 3379–3394.
- Beyer HM, Gonschorek P, Samodelov SL, Meier M, Weber W, Zurbriggen MD.** 2015. AQUA Cloning: A Versatile and Simple Enzyme-Free Cloning Approach. *PLOS ONE* **10**, e0137652.
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U.** 2009. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509–1512.

- Boersma MR, Patrick RM, Jillings SL, Shaipulah NFM, Sun P, Haring MA, Dudareva N, Li Y, Schuurink RC.** 2021. ODORANT1 targets multiple metabolic networks in petunia flowers. *The Plant Journal* **n/a**.
- Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.
- Bortesi L, Rademacher T, Schiermeyer A, Schuster F, Pezzotti M, Schillberg S.** 2012. Development of an optimized tetracycline-inducible expression system to increase the accumulation of interleukin-10 in tobacco BY-2 suspension cells. *BMC biotechnology* **12**, 40–40.
- Bothfeld W, Kapov G, Tyo KEJ.** 2017. A Glucose-Sensing Toggle Switch for Autonomous, High Productivity Genetic Control. *ACS Synthetic Biology* **6**, 1296–1304.
- Boyle PM, Burrill DR, Inniss MC, et al.** 2012. A BioBrick compatible strategy for genetic modification of plants. *Journal of Biological Engineering* **6**, 8.
- Breitling R, Takano E.** 2015. Synthetic biology advances for pharmaceutical production. *Chemical biotechnology • Pharmaceutical biotechnology* **35**, 46–51.
- Bruce W, Folkerts O, Garnaat C, Crasta O, Roth B, Bowen B.** 2000. Expression Profiling of the Maize Flavonoid Pathway Genes Controlled by Estradiol-Inducible Transcription Factors CRC and P. *The Plant Cell* **12**, 65–79.
- Brückner K, Schäfer P, Weber E, Grützner R, Marillonnet S, Tissier A.** 2015. A library of synthetic transcription activator-like effector-activated promoters for coordinated orthogonal gene expression in plants. *The Plant Journal* **82**, 707–716.



- Butelli E, Titta L, Giorgio M, et al.** 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nature Biotechnology* **26**, 1301–1308.
- Caddick MX, Greenland AJ, Jepson Ian, Krause K-P, Qu N, Riddell KV, Salter MG, Schuch W, Sonnewald U, Tomsett AB.** 1998. An ethanol inducible gene switch for plants used to manipulate carbon metabolism. *Nature Biotechnology* **16**, 177–180.
- Cameron DE, Bashor CJ, Collins JJ.** 2014. A brief history of synthetic biology. *Nature Reviews Microbiology* **12**, 381–390.
- Campbell MEM, Palfreyman JW, Preston CM.** 1984. Identification of herpes simplex virus DNA sequences which encode a trans-acting polypeptide responsible for stimulation of immediate early transcription. *Journal of Molecular Biology* **180**, 1–19.
- Carbonell P, Jervis AJ, Robinson CJ, et al.** 2018. An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Communications Biology* **1**, 66.
- Carey M, Leatherwood J, Ptashne M.** 1990. A Potent GAL4 Derivative Activates Transcription at a Distance in Vitro. *Science* **247**, 710–712.
- Chaiwanon J, Wang Z-Y.** 2015. Spatiotemporal Brassinosteroid Signaling and Antagonism with Auxin Pattern Stem Cell Dynamics in Arabidopsis Roots. *Current Biology* **25**, 1031–1042.

- Charfeddine M, Samet M, Charfeddine S, Bouaziz D, Gargouri Bouzid R.** 2019. Ectopic Expression of StERF94 Transcription Factor in Potato Plants Improved Resistance to *Fusarium solani* Infection. *Plant Molecular Biology Reporter* **37**, 450–463.
- Chatelle C, Ochoa-Fernandez R, Engesser R, Schneider N, Beyer HM, Jones AR, Timmer J, Zurbriggen MD, Weber W.** 2018. A Green-Light-Responsive System for the Control of Transgene Expression in Mammalian and Plant Cells. *ACS Synthetic Biology* **7**, 1349–1358.
- Chavez A, Scheiman J, Vora S, et al.** 2015. Highly efficient Cas9-mediated transcriptional programming. *Nature methods* **12**, 326–328.
- Chen L, Aihara K.** 2002. Stability of genetic regulatory networks with time delay. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* **49**, 602–608.
- Chen J, Clinton M, Qi G, Wang D, Liu F, Fu ZQ.** 2020. Reprogramming and remodeling: transcriptional and epigenetic regulation of salicylic acid-mediated plant defense. *Journal of Experimental Botany* **71**, 5256–5268.
- Chen H, Kazemier HG, de Groote ML, Ruiters MHJ, Xu G-L, Rots MG.** 2014. Induced DNA demethylation by targeting Ten-Eleven Translocation 2 to the human ICAM-1 promoter. *Nucleic Acids Research* **42**, 1563–1574.
- Cheng H, Song S, Xiao L, Soo HM, Cheng Z, Xie D, Peng J.** 2009. Gibberellin Acts through Jasmonate to Control the Expression of MYB21, MYB24, and MYB57 to Promote Stamen Filament Growth in *Arabidopsis*. *PLOS Genetics* **5**, e1000440.

- Chini A, Fonseca S, Fernández G, et al.** 2007. The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **448**, 666–671.
- Cho SW, Kim S, Kim JM, Kim J-S.** 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology* **31**, 230–232.
- Christie JM, Zurbriggen MD.** 2021. Optogenetics in plants. *New Phytologist* **229**, 3108–3115.
- Clemente T.** 2006. *Nicotiana (Nicotiana tabaccum, Nicotiana benthamiana)*. In: Wang K, ed. *Agrobacterium protocols*. Springer, 143–154.
- Cody WB, Scholthof HB, Mirkov TE.** 2017. Multiplexed Gene Editing and Protein Overexpression Using a Tobacco mosaic virus Viral Vector. *Plant Physiology* **175**, 23–35.
- Cong L, Ran FA, Cox D, et al.** 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823.
- Cruz-Ramírez A, Díaz-Triviño S, Blilou I, et al.** 2012. A Bistable Circuit Involving SCARECROW-RETINOBLASTOMA Integrates Cues to Inform Asymmetric Stem Cell Division. *Cell* **150**, 1002–1015.
- D’Agostino IB, Deruère J, Kieber JJ.** 2000. Characterization of the Response of the Arabidopsis Response Regulator Gene Family to Cytokinin1. *Plant Physiology* **124**, 1706–1717.
- Dal Cin V, Tieman DM, Tohge T, et al.** 2011. Identification of genes in the phenylalanine metabolic pathway by ectopic expression of a MYB transcription factor in tomato fruit. *The Plant cell* **23**, 2738–2753.

- Danilo B, Perrot L, Mara K, Botton E, Nogué F, Mazier M.** 2019. Efficient and transgene-free gene targeting using *Agrobacterium*-mediated delivery of the CRISPR/Cas9 system in tomato. *Plant Cell Reports* **38**, 459–462.
- D’Aoust M-A, Lavoie P-O, Couture MM-J, Trépanier S, Guay J-M, Dargis M, Mongrand S, Landry N, Ward BJ, Vézina L-P.** 2008. Influenza virus-like particles produced by transient expression in *Nicotiana benthamiana* induce a protective immune response against a lethal viral challenge in mice. *Plant Biotechnology Journal* **6**, 930–940.
- Davies HM.** 2010. Review article: Commercialization of whole-plant systems for biomanufacturing of protein products: evolution and prospects. *Plant Biotechnology Journal* **8**, 845–861.
- De Las Heras A, Carreño CA, Martínez-García E, De Lorenzo V.** 2010. Engineering input/output nodes in prokaryotic regulatory circuits. *FEMS Microbiology Reviews* **34**, 842–865.
- Deans TL, Cantor CR, Collins JJ.** 2007. A Tunable Genetic Switch Based on RNAi and Repressor Proteins for Regulating Gene Expression in Mammalian Cells. *Cell* **130**, 363–372.
- Dias AP, Grotewold E.** 2003. Manipulating the accumulation of phenolics in maize cultured cells using transcription factors. *Advance in Plant Anthocyanin Research and Development* **14**, 207–216.

- Dickmeis C, Fischer R, Commandeur U.** 2014. Potato virus X-based expression vectors are stabilized for long-term production of proteins and larger inserts. *Biotechnology Journal* **9**, 1369–1379.
- Diego-Martin B, González B, Vazquez-Vilar M, Selma S, Mateos-Fernández R, Gianoglio S, Fernández-del-Carmen A, Orzáez D.** 2020. Pilot Production of SARS-CoV-2 Related Proteins in Plants: A Proof of Concept for Rapid Repurposing of Indoor Farms Into Biomanufacturing Facilities. *Frontiers in Plant Science* **11**, 2101.
- Dilão R.** 2014. The regulation of gene expression in eukaryotes: Bistability and oscillations in repressilator models. *Journal of Theoretical Biology* **340**, 199–208.
- Dill GM.** 2005. Glyphosate-resistant crops: history, status and future. *Pest Management Science* **61**, 219–224.
- Din MO, Danino T, Prindle A, et al.** 2016. Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* **536**, 81–85.
- Dong H, Liu L, Fan X, et al.** 2019. The Artificial Promoter rMdAG2I Confers Flower-specific Activity in Malus. *International Journal of Molecular Sciences* **20**.
- Duan L, Hope J, Ong Q, Lou H-Y, Kim N, McCarthy C, Acero V, Lin MZ, Cui B.** 2017. Understanding CRY2 interactions for optical control of intracellular signaling. *Nature Communications* **8**, 547.
- Durai S, Mani M, Kandavelou K, Wu J, Porteus MH, Chandrasegaran S.** 2005. Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Research* **33**, 5978–5990.

- Ede C, Chen X, Lin M-Y, Chen YY.** 2016. Quantitative Analyses of Core Promoters Enable Precise Engineering of Regulated Gene Expression in Mammalian Cells. *ACS synthetic biology* **5**, 395–404.
- Ellison EE, Nagalakshmi U, Gamo ME, Huang P, Dinesh-Kumar S, Voytas DF.** 2020. Multiplexed heritable gene editing using RNA viruses and mobile single guide RNAs. *Nature Plants* **6**, 620–624.
- Elowitz MB, Leibler S.** 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338.
- Estruch JJ, Carozzi NB, Desai N, Duck NB, Warren GW, Koziel MG.** 1997. Transgenic plants: An emerging approach to pest control. *Nature Biotechnology* **15**, 137–141.
- Fang C, Fernie AR, Luo J.** 2019. Exploring the Diversity of Plant Metabolism. *Trends in Plant Science* **24**, 83–98.
- Farzadfard F, Perli SD, Lu TK.** 2013. Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. *ACS Synthetic Biology* **2**, 604–613.
- Fresquet-Corrales S, Roque E, Sarrión-Perdigones A, Rochina M, López-Gresa MP, Díaz-Mula HM, Bellés JM, Tomás-Barberán F, Beltrán JP, Cañas LA.** 2017. Metabolic engineering to simultaneously activate anthocyanin and proanthocyanidin biosynthetic pathways in *Nicotiana* spp. *PLOS ONE* **12**, e0184839.
- Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ.** 2009. Synthetic Gene Networks That Count. *Science* **324**, 1199–1202.
- Gallego-Bartolomé J.** 2020. DNA methylation in plants: mechanisms and tools for targeted manipulation. *New Phytologist* **227**, 38–44.

- Gallego-Bartolomé J, Gardiner J, Liu W, Papikian A, Ghoshal B, Kuo HY, Zhao JM-C, Segal DJ, Jacobsen SE.** 2018. Targeted DNA demethylation of the Arabidopsis genome using the human TET1 catalytic domain. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E2125–E2134.
- Gamir J, Pastor V, Cerezo M, Flors V.** 2012. Identification of indole-3-carboxylic acid as mediator of priming against *Plectosphaerella cucumerina*. *Plant physiology and biochemistry : PPB* **61**, 169–179.
- Gao C.** 2019. Precision plant breeding using genome editing technologies. *Transgenic Research* **28**, 53–55.
- Gao X, Tsang JCH, Gaba F, Wu D, Lu L, Liu P.** 2014. Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic acids research* **42**, e155.
- Garcia-Ojalvo J, Elowitz MB, Strogatz SH.** 2004. Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10955.
- Garcia-Perez E, Diego-Martin B, Quijano-Rubio A, Moreno-Gimenez E, Orzaez D, Vazquez-Vilar M.** 2021. A copper switch for inducing CRISPR/Cas9-based transcriptional activation tightly regulates gene expression in *Nicotiana benthamiana*. *bioRxiv*, 2021.09.07.459151.
- Garcia-Perez E, Diego-Martin B, Quijano-Rubio A, Moreno-Giménez E, Selma S, Orzaez D, Vazquez-Vilar M.** 2022. A copper switch for inducing CRISPR/Cas9-based

- transcriptional activation tightly regulates gene expression in *Nicotiana benthamiana*. *BMC Biotechnology* **22**, 12.
- Gardner TS, Cantor CR, Collins JJ.** 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342.
- Gatz C, Froberg C, Wendenburg R.** 1992. Stringent repression and homogeneous de-repression by tetracycline of a modified CaMV 35S promoter in intact transgenic tobacco plants. *The Plant Journal* **2**, 397–404.
- Gentzel IN, Ohlson EW, Redinbaugh MG, Wang G-L.** 2022. VIGE: virus-induced genome editing for improving abiotic and biotic stress traits in plants. *Stress Biology* **2**, 2.
- Georgianna DR, Mayfield SP.** 2012. Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature* **488**, 329–335.
- Ghoshal B, Vong B, Picard CL, Feng S, Tam JM, Jacobsen SE.** 2020. A viral guide RNA delivery system for CRISPR-based transcriptional activation and heritable targeted DNA demethylation in *Arabidopsis thaliana*. *PLoS genetics* **16**, e1008983–e1008983.
- Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO.** 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345.
- Gleba YY, Tusé D, Giritch A.** 2013. Plant Viral Vectors for Delivery by *Agrobacterium*. In: Palmer K,, In: Gleba Y, eds. *Current Topics in Microbiology and Immunology. Plant Viral Vectors*. Berlin, Heidelberg: Springer Berlin Heidelberg, 155–192.
- Godwin ID, Rutkoski J, Varshney RK, Hickey LT.** 2019. Technological perspectives for plant breeding. *Theoretical and Applied Genetics* **132**, 555–557.



- Goñi-Moreno A, Amos M.** 2012. A reconfigurable NAND/NOR genetic logic gate. *BMC Systems Biology* **6**, 126.
- Gosal SS, Wani SH, Kang MS.** 2010. Biotechnology and Crop Improvement. *Journal of Crop Improvement* **24**, 153–217.
- Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A.** 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research* **36**, 3420–3435.
- Grosse-Holz F, Madeira L, Zahid MA, Songer M, Kourelis J, Fesenko M, Ninck S, Kaschani F, Kaiser M, van der Hoorn RAL.** 2018. Three unrelated protease inhibitors enhance accumulation of pharmaceutical recombinant proteins in *Nicotiana benthamiana*. *Plant Biotechnology Journal* **16**, 1797–1810.
- Guan X, Stege J, Kim M, Dahmani Z, Fan N, Heifetz P, Barbas CF, Briggs SP.** 2002. Heritable endogenous gene regulation in plants with designed polydactyl zinc finger transcription factors. *Proceedings of the National Academy of Sciences* **99**, 13296–13301.
- Guo H, Ecker JR.** 2003. Plant Responses to Ethylene Gas Are Mediated by SCFEBF1/EBF2-Dependent Proteolysis of EIN3 Transcription Factor. *Cell* **115**, 667–677.
- Haeussler M, Schönig K, Eckert H, et al.** 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology* **17**, 148.
- Hamada H, Linghu Q, Nagira Y, Miki R, Taoka N, Imai R.** 2017. An in planta biolistic method for stable wheat transformation. *Scientific Reports* **7**, 11443.

- Haran S, Logendra S, Seskar M, Bratanova M, Raskin I.** 2000. Characterization of Arabidopsis acid phosphatase promoter and regulation of acid phosphatase expression. *Plant Physiol* **124** : **2**, 615–626.
- Haseloff J, Ajioka J.** 2009. Synthetic biology: history, challenges and prospects. *Journal of The Royal Society Interface* **6**, S389–S391.
- Haseloff J, Siemering KR.** 2006. The uses of green fluorescent protein in plants. *Methods of biochemical analysis* **47**, 259–284.
- Hashimoto R, Ueta R, Abe C, Osakabe Y, Osakabe K.** 2018. Efficient Multiplex Genome Editing Induces Precise, and Self-Ligated Type Mutations in Tomato Plants. *Frontiers in Plant Science* **9**, 916.
- Havens KA, Guseman JM, Jang SS, Pierre-Jerome E, Bolten N, Klavins E, Nemhauser JL.** 2012. A Synthetic Approach Reveals Extensive Tunability of Auxin Signaling. *Plant Physiology* **160**, 135–142.
- van Herpen TWJM, Cankar K, Nogueira M, Bosch D, Bouwmeester HJ, Beekwilder J.** 2010. *Nicotiana benthamiana* as a Production Platform for Artemisinin Precursors. *PLOS ONE* **5**, e14222.
- Herud-Sikimić O, Stiel AC, Kolb M, Shanmugaratnam S, Berendzen KW, Feldhaus C, Höcker B, Jürgens G.** 2021. A biosensor for the direct visualization of auxin. *Nature* **592**, 768–772.
- Hicks M, Bachmann TT, Wang B.** 2020. Synthetic Biology Enables Programmable Cell-Based Biosensors. *ChemPhysChem* **21**, 132–144.

- Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA.** 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology* **33**, 510–517.
- Hiratsu K, Matsui K, Koyama T, Ohme-Takagi M.** 2003. Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *The Plant Journal* **34**, 733–739.
- Ho H-I, Fang JR, Cheung J, Wang HH.** 2020. Programmable CRISPR-Cas transcriptional activation in bacteria. *Molecular Systems Biology* **16**, e9427.
- Hobom B.** 1980. Surgery of genes-at the doorstep of synthetic biology. *Medizinische Klinik* **75**, 14–21.
- Hong JC.** 2016. Chapter 3 - General Aspects of Plant Transcription Factor Families. In: Gonzalez DH, ed. *Plant Transcription Factors*. Boston: Academic Press, 35–56.
- Hope IA, Struhl K.** 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of Yeast. *Cell* **46**, 885–894.
- Hu J, Li S, Li Z, Li H, Song W, Zhao H, Lai J, Xia L, Li D, Zhang Y.** 2019. A barley stripe mosaic virus-based guide RNA delivery system for targeted mutagenesis in wheat and maize. *Molecular Plant Pathology* **20**, 1463–1474.
- Huang H, Gong Y, Liu B, Wu D, Zhang M, Xie D, Song S.** 2020. The DELLA proteins interact with MYB21 and MYB24 to regulate filament elongation in *Arabidopsis*. *BMC Plant Biology* **20**, 64.
- Huang D, Kosentka PZ, Liu W.** 2021. Synthetic biology approaches in regulation of targeted gene expression. *Cell signaling and gene regulation* **63**, 102036.

- Ikeda M, Ohme-Takagi M.** 2009. A Novel Group of Transcriptional Repressors in Arabidopsis. *Plant and Cell Physiology* **50**, 970–975.
- Iverson SV, Haddock TL, Beal J, Densmore DM.** 2016. CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synthetic Biology* **5**, 99–103.
- Jiang W, Li H, Liu X, Zhang J, Zhang W, Li T, Liu L, Yu X.** 2020. Precise and efficient silencing of mutant Kras(G12D) by CRISPR-CasRx controls pancreatic cancer progression. *Theranostics* **10**, 11507–11519.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E.** 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)* **337**, 816–821.
- Jisha V, Dampanaboina L, Vadassery J, Mithöfer A, Kappara S, Ramanan R.** 2015. Overexpression of an AP2/ERF Type Transcription Factor OsEREBP1 Confers Biotic and Abiotic Stress Tolerance in Rice. *PLOS ONE* **10**, e0127831.
- Jones AM, Danielson JA, Manojkumar SN, Lanquar V, Grossmann G, Frommer WB.** 2014. Abscisic acid dynamics in roots detected with genetically encoded FRET sensors. *eLife* **3**, e01741–e01741.
- Kagale S, Rozwadowski K.** 2011. EAR motif-mediated transcriptional repression in plants. *Epigenetics* **6**, 141–146.
- Kassaw TK, Donayre-Torres AJ, Antunes MS, Morey KJ, Medford JI.** 2018. Engineering synthetic regulatory circuits in plants. *Synthetic Biology Meets Plant Metabolism* **273**, 13–22.

- Keating KW, Young EM.** 2019. Synthetic biology for bio-derived structural materials. *Materials engineering: bio-derived/bio-inspired materials • Separations Engineering: advances in adsorption* **24**, 107–114.
- Keegan L, Gill G, Ptashne M.** 1986. Separation of DNA Binding from the Transcription-Activating Function of a Eukaryotic Regulatory Protein. *Science* **231**, 699–704.
- Khakhar A, Wang C, Swanson R, Stokke S, Rizvi F, Sarup S, Hobbs J, Voytas DF.** 2021. VipariNama: RNA viral vectors to rapidly elucidate the relationship between gene expression and phenotype. *Plant Physiology* **186**, 2222–2238.
- Khalil AS, Collins JJ.** 2010. Synthetic biology: applications come of age. *Nature Reviews Genetics* **11**, 367–379.
- Khan MK, Zill-E-Huma, Dangles O.** 2014. A comprehensive review on flavanones, the major citrus polyphenols. *Journal of Food Composition and Analysis* **33**, 85–104.
- Kim D, Langmead B, Salzberg SL.** 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357–360.
- Kim D-H, Yu J, Park JC, Jeong C-B, Bae S, Lee J-S.** 2019. Targeted cytochrome P450 3045C1 (CYP3045C1) gene mutation via CRISPR-Cas9 ribonucleoproteins in the marine rotifer *Brachionus koreanus*. *Hydrobiologia* **844**, 117–128.
- Klein J, Heal JR, Hamilton WDO, Boussemerghoune T, Tange TØ, Delegrange F, Jaeschke G, Hatsch A, Heim J.** 2014. Yeast Synthetic Biology Platform Generates Novel Chemical Structures as Scaffolds for Drug Discovery. *ACS Synthetic Biology* **3**, 314–323.

- Koide T, Lee Pang W, Baliga NS.** 2009. The role of predictive modelling in rationally re-engineering biological systems. *Nature Reviews Microbiology* **7**, 297–305.
- Konermann S, Brigham MD, Trevino AE, et al.** 2015*a*. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*.
- Konermann S, Brigham MD, Trevino AE, et al.** 2015*b*. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*.
- Koo JC, Asurmendi S, Bick J, Woodford-Thomas T, Beachy RN.** 2004. Ecdysone agonist-inducible expression of a coat protein gene from tobacco mosaic virus confers viral resistance in transgenic Arabidopsis. *The Plant Journal* **37**, 439–448.
- Koonin EV, Makarova KS.** 2019. Origins and evolution of CRISPR-Cas systems. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **374**, 20180087–20180087.
- Kumar S, AlAbed D, Witteck JT, et al.** 2015. A combinatorial bidirectional and bicistronic approach for coordinated multi-gene expression in corn. *Plant Molecular Biology* **87**, 341–353.
- Larrieu A, Champion A, Legrand J, et al.** 2015. A fluorescent hormone biosensor reveals the dynamics of jasmonate signalling in plants. *Nature Communications* **6**, 6043.
- Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS.** 2013. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols* **8**, 2180–2196.

- Laughon A, Gesteland RF.** 1982. Isolation and preliminary characterization of the GAL4 gene, a positive regulator of transcription in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 6827–6831.
- Lee JE, Neumann M, Duro DI, Schmid M.** 2019. CRISPR-based tools for targeted transcriptional and epigenetic regulation in plants. *PLOS ONE* **14**, e0222778.
- Lee M, Woo HM.** 2020. A Logic NAND Gate for Controlling Gene Expression in a Circadian Rhythm in Cyanobacteria. *ACS Synthetic Biology* **9**, 3210–3216.
- Lessard PA, Kulaveerasingam H, York GM, Strong A, Sinskey AJ.** 2002. Manipulating Gene Expression for the Metabolic Engineering of Plants. *Metabolic Engineering* **4**, 67–79.
- Leydon AR, Gala HP, Guiziou S, Nemhauser JL.** 2020. Engineering Synthetic Signaling in Plants. *Annual Review of Plant Biology* **71**, 767–788.
- Li J, Blue R, Zeitler B, Strange TL, Pearl JR, Huizinga DH, Evans S, Gregory PD, Urnov FD, Petolino JF.** 2013. Activation domains for controlling plant gene expression using designed transcription factors. *Plant biotechnology journal* **11**, 671–680.
- Li T, Hu J, Sun Y, *et al.*** 2021. Highly efficient heritable genome editing in wheat using an RNA virus and bypassing tissue culture. *Molecular Plant* **14**, 1787–1798.
- Li R, Jia X, Mao X.** 2005. Ethanol-inducible gene expression system and its applications in plant functional genomics. *Plant Science* **169**, 463–469.
- Li W, Ma M, Feng Y, Li H, Wang Y, Ma Y, Li M, An F, Guo H.** 2015*a*. EIN2-Directed Translational Regulation of Ethylene Signaling in Arabidopsis. *Cell* **163**, 670–683.

- Li Y, Pearl SA, Jackson SA.** 2015*b*. Gene Networks in Plant Biology: Approaches in Reconstruction and Analysis. *Trends in Plant Science* **20**, 664–675.
- Li Z, Wang F, Li J-F.** 2019. Targeted Transcriptional Activation in Plants Using a Potent Dead Cas9-Derived Synthetic Gene Activator. *Current protocols in molecular biology* **127**, e89–e89.
- Li Z, Zhang D, Xiong X, Yan B, Xie W, Sheen J, Li JF.** 2017*a*. A potent Cas9-derived gene activator for plant and mammalian cells. *Nature Plants* **3**, 930–936.
- Li Z, Zhang D, Xiong X, Yan B, Xie W, Sheen J, Li J-F.** 2017*b*. A potent Cas9-derived gene activator for plant and mammalian cells. *Nature Plants* **3**, 930–936.
- Liang Z, Chen K, Gao C.** 2019. Biolistic Delivery of CRISPR/Cas9 with Ribonucleoprotein Complex in Wheat. In: Qi Y, ed. *Plant Genome Editing with CRISPR Systems: Methods and Protocols*. New York, NY: Springer New York, 327–335.
- Liao C-Y, Smet W, Brunoud G, Yoshida S, Vernoux T, Weijers D.** 2015. Reporters for sensitive and quantitative measurement of auxin response. *Nature Methods* **12**, 207–210.
- Lico C, Benvenuto E, Baschieri S.** 2015. The Two-Faced Potato Virus X: From Plant Pathogen to Smart Nanoparticle. *Frontiers in Plant Science* **6**.
- Lieberman-Lazarovich M, Yahav C, Israeli A, Efroni I.** 2019. Deep Conservation of cis-Element Variants Regulating Plant Hormonal Responses. *The Plant Cell* **31**, 2559–2572.



- Lindhout BI, Pinas JE, Hooykaas PJJ, Van Der Zaal BJ.** 2006. Employing libraries of zinc finger artificial transcription factors to screen for homologous recombination mutants in Arabidopsis. *The Plant Journal* **48**, 475–483.
- Liu J, Osbourn A, Ma P.** 2015. MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. *Molecular Plant* **8**, 689–708.
- Liu W, Rudis MR, Peng Y, Mazarei M, Millwood RJ, Yang J-P, Xu W, Chesnut JD, Stewart Jr CN.** 2014. Synthetic TAL effectors for targeted enhancement of transgene expression in plants. *Plant Biotechnology Journal* **12**, 436–446.
- Liu D, Shi L, Han C, Yu J, Li D, Zhang Y.** 2012. Validation of Reference Genes for Gene Expression Studies in Virus-Infected *Nicotiana benthamiana* Using Quantitative Real-Time PCR. *PLOS ONE* **7**, e46451.
- Liu W, Stewart CN.** 2015. Plant synthetic biology. *Trends in Plant Science* **20**, 309–317.
- Liu W, Yuan JS, Stewart Jr CN.** 2013. Advanced genetic tools for plant biotechnology. *Nature Reviews Genetics* **14**, 781–793.
- Liu Z, Zhang Y, Nielsen J.** 2019. Synthetic Biology of Yeast. *Biochemistry* **58**, 1511–1520.
- Livak KJ, Schmittgen TD.** 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods* **25**, 402–408.
- Llorente B, Torres-Montilla S, Morelli L, et al.** 2020. Synthetic conversion of leaf chloroplasts into carotenoid-rich plastids reveals mechanistic basis of natural chromoplast development. *Proceedings of the National Academy of Sciences* **117**, 21796 LP – 21803.

- Louveau T, Osbourn A.** 2019. The Sweet Side of Plant-Specialized Metabolism. *Cold Spring Harbor perspectives in biology* **11**, a034744.
- Lowder LG, Paul JW, Qi Y.** 2017. Multiplexed Transcriptional Activation or Repression in Plants Using CRISPR-dCas9-Based Systems. In: Kaufmann K, In: Mueller-Roeber B, eds. *Plant Gene Regulatory Networks: Methods and Protocols*. New York, NY: Springer New York, 167–184.
- Lowder LG, Zhang D, Baltes NJ, Paul JW 3rd, Tang X, Zheng X, Voytas DF, Hsieh T-F, Zhang Y, Qi Y.** 2015. A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation. *Plant physiology* **169**, 971–985.
- Lowder LG, Zhou J, Zhang Y, Malzahn A, Zhong Z, Hsieh T-F, Voytas DF, Zhang Y, Qi Y.** 2018*a*. Robust Transcriptional Activation in Plants Using Multiplexed CRISPR-Act2.0 and mTALE-Act Systems. *Molecular Plant* **11**, 245–256.
- Lowder LG, Zhou J, Zhang Y, Malzahn A, Zhong Z, Hsieh T-F, Voytas DF, Zhang Y, Qi Y.** 2018*b*. Robust Transcriptional Activation in Plants Using Multiplexed CRISPR-Act2.0 and mTALE-Act Systems. *Molecular Plant* **11**, 245–256.
- Lu R, Martin-Hernandez AM, Peart JR, Malcuit I, Baulcombe DC.** 2003. Virus-induced gene silencing in plants. *RNA interference* **30**, 296–303.
- Lucks JB, Qi L, Whitaker WR, Arkin AP.** 2008. Toward scalable parts families for predictable design of biological circuits. *Growth and Development: Eukaryotes/Prokaryotes* **11**, 567–573.
- Ma JK-C, Drake PMW, Christou P.** 2003. The production of recombinant pharmaceutical proteins in plants. *Nature Reviews Genetics* **4**, 794–805.

- Maeda AE, Nakamichi N.** 2022. Plant clock modifications for adapting flowering time to local environments. *Plant Physiology*, kiac107.
- Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK.** 2013. CRISPR RNA-guided activation of endogenous human genes. *Nature Methods* **10**, 977–979.
- Mahas A, Aman R, Mahfouz M.** 2019. CRISPR-Cas13d mediates robust RNA virus interference in plants. *Genome biology* **20**, 263–263.
- Maioli A, Gianoglio S, Moglia A, et al.** 2020. Simultaneous CRISPR/Cas9 Editing of Three PPO Genes Reduces Fruit Flesh Browning in *Solanum melongena* L. *Frontiers in Plant Science* **11**, 1883.
- Mali P, Esvelt KM, Church GM.** 2013. Cas9 as a versatile tool for engineering biology. *Nature Methods* **10**, 957–963.
- Mao Y, Botella JR, Liu Y, Zhu J-K.** 2019. Gene editing in plants: progress and challenges. *National Science Review* **6**, 421–437.
- Marillonnet S, Thoeringer C, Kandzia R, Klimyuk V, Gleba Y.** 2005. Systemic *Agrobacterium tumefaciens*-mediated transfection of viral replicons for efficient transient expression in plants. *Nature Biotechnology* **23**, 718–723.
- Mathews H, Clendennen SK, Caldwell CG, et al.** 2003. Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. *The Plant cell* **15**, 1689–1703.
- Mathur J, Koncz C.** 1998. PEG-mediated protoplast transformation with naked DNA. *Methods in molecular biology* (Clifton, N.J.) **82**, 267–276.

- McCarthy DM, Medford JI.** 2020. Quantitative and Predictive Genetic Parts for Plant Synthetic Biology. *Frontiers in Plant Science* **11**.
- McKenzie MJ, Mett VV, PH SR, Jameson PE.** 1998. Controlled cytokinin production in transgenic tobacco using a copper-inducible promoter. *Plant physiology* **116**, 969–977.
- Mei Y, Beernink BM, Ellison EE, Konečná E, Neelakandan AK, Voytas DF, Whitham SA.** 2019. Protein expression and gene editing in monocots using foxtail mosaic virus vectors. *Plant direct* **3**, e00181–e00181.
- Merchante C, Brumos J, Yun J, Hu Q, Spencer KR, Enríquez P, Binder BM, Heber S, Stepanova AN, Alonso JM.** 2015. Gene-Specific Translation Regulation Mediated by the Hormone-Signaling Molecule EIN2. *Cell* **163**, 684–697.
- Misra P, Pandey A, Tiwari M, et al.** 2010. Modulation of transcriptome and metabolome of tobacco by Arabidopsis transcription factor, AtMYB12, leads to insect resistance. *Plant physiology* **152**, 2258–2268.
- Mohan C, Jayanarayanan AN, Narayanan S.** 2017. Construction of a novel synthetic root-specific promoter and its characterization in transgenic tobacco plants. *3 Biotech* **7**, 234.
- Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E.** 2005. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution* **60**, 174–182.

- Molina-Hidalgo FJ, Vazquez-Vilar M, D'Andrea L, Demurtas OC, Fraser P, Giuliano G, Bock R, Orzáez D, Goossens A.** 2020. Engineering Metabolism in Nicotiana Species: A Promising Future. *Trends in biotechnology*.
- Moore R, Chandrahas A, Bleris L.** 2014. Transcription Activator-like Effectors: A Toolkit for Synthetic Biology. *ACS Synthetic Biology* **3**, 708–716.
- Moore I, Gälweiler L, Grosskopf D, Schell J, Palme K.** 1998. A transcription activation system for regulated gene expression in transgenic plants. *Proceedings of the National Academy of Sciences* **95**, 376–381.
- Moradpour M, Abdulah SNA.** 2020. CRISPR/dCas9 platforms in plants: strategies and applications beyond genome editing. *Plant Biotechnology Journal* **18**, 32–44.
- Morbitzer R, Römer P, Boch J, Lahaye T.** 2010. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21617–21622.
- Mou Z, Fan W, Dong X.** 2003. Inducers of Plant Systemic Acquired Resistance Regulate NPR1 Function through Redox Changes. *Cell* **113**, 935–944.
- Mózsik L, Hoekzema M, de Kok NAW, Bovenberg RAL, Nygård Y, Driessen AJM.** 2021. CRISPR-based transcriptional activation tool for silent genes in filamentous fungi. *Scientific Reports* **11**, 1118.
- Müller K, Siegel D, Rodriguez Jahnke F, Gerrer K, Wend S, Decker EL, Reski R, Weber W, Zurbriggen MD.** 2014. A red light-controlled synthetic gene expression switch for plant systems. *Molecular BioSystems* **10**, 1679–1688.

- Nabavi SM, Šamec D, Tomczyk M, et al.** 2020. Flavonoid biosynthetic pathways in plants: Versatile targets for metabolic engineering. *Biotechnology Advances* **38**, 107316–107316.
- Naing AH, Park KI, Ai TN, Chung MY, Han JS, Kang Y-W, Lim KB, Kim CK.** 2017. Overexpression of snapdragon Delila (Del) gene in tobacco enhances anthocyanin accumulation and abiotic stress tolerance. *BMC plant biology* **17**, 65–65.
- Neelam, Khatkar A, Sharma KK.** 2020. Phenylpropanoids and its derivatives: biological activities and its role in food, pharmaceutical and cosmetic industries. *Critical Reviews in Food Science and Nutrition* **60**, 2655–2675.
- Norkunas K, Harding R, Dale J, Dugdale B.** 2018. Improving agroinfiltration-based transient gene expression in *Nicotiana benthamiana*. *Plant Methods* **14**, 71.
- Ochoa-Fernandez R, Abel NB, Wieland F-G, et al.** 2020. Optogenetic control of gene expression in plants in the presence of ambient white light. *Nature Methods* **17**, 717–725.
- Ochoa-Fernandez R, Samodelov SL, Brandl SM, Wehinger E, Müller K, Weber W, Zurbriggen MD.** 2016. Optogenetics in Plants: Red/Far-Red Light Control of Gene Expression. *Methods in molecular biology (Clifton, N.J.)* **1408**, 125–139.
- Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-Takagi M.** 2001. Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *The Plant cell* **13**, 1959–1968.
- Okada R, Nemoto Y, Endo-Higashi N, Izawa T.** 2017. Synthetic control of flowering in rice independent of the cultivation environment. *Nature Plants* **3**, 17039.

- Okuzaki A, Konagaya K, Nanasato Y, Tsuda M, Tabei Y.** 2011. Estrogen-inducible GFP expression patterns in rice (*Oryza sativa* L.). *Plant Cell Reports* **30**, 529–538.
- Ordiz MI, Barbas CF, Beachy RN.** 2002. Regulation of transgene expression in plants with polydactyl zinc finger transcription factors. *Proceedings of the National Academy of Sciences* **99**, 13290–13295.
- Outchkourov NS, Carollo CA, Gomez-Roldan V, de Vos RCH, Bosch D, Hall RD, Beekwilder J.** 2014. Control of anthocyanin and non-flavonoid compounds by anthocyanin-regulating MYB and bHLH transcription factors in *Nicotiana benthamiana* leaves. *Frontiers in Plant Science* **5**, 1–9.
- Pan C, Sretenovic S, Qi Y.** 2021a. CRISPR/dCas-mediated transcriptional and epigenetic regulation in plants. *Current Opinion in Plant Biology* **60**, 101980.
- Pan C, Wu X, Markel K, Malzahn AA, Kundagrami N, Sretenovic S, Zhang Y, Cheng Y, Shih PM, Qi Y.** 2021b. CRISPR–Act3.0 for highly efficient multiplexed gene activation in plants. *Nature Plants*.
- Pan C, Wu X, Markel K, Malzahn AA, Kundagrami N, Sretenovic S, Zhang Y, Cheng Y, Shih PM, Qi Y.** 2021c. CRISPR–Act3.0 for highly efficient multiplexed gene activation in plants. *Nature Plants* **7**, 942–953.
- Papanatsiou M, Petersen J, Henderson L, Wang Y, Christie JM, Blatt MR.** 2019. Optogenetic manipulation of stomatal kinetics improves carbon assimilation, water use, and growth. *Science (New York, N.Y.)* **363**, 1456–1459.

- Papikian A, Liu W, Gallego-Bartolomé J, Jacobsen SE.** 2019. Site-specific manipulation of Arabidopsis loci using CRISPR-Cas9 SunTag systems. *Nature communications* **10**, 729.
- Park J-J, Dempewolf E, Zhang W, Wang Z-Y.** 2017. RNA-guided transcriptional activation via CRISPR/dCas9 mimics overexpression phenotypes in Arabidopsis. *PLOS ONE* **12**, e0179410.
- Park S-Y, Peterson FC, Mosquna A, Yao J, Volkman BF, Cutler SR.** 2015. Agrochemical control of plant water use using engineered abscisic acid receptors. *Nature* **520**, 545–548.
- Park CH, Xu H, Yeo HJ, Park YE, Hwang G-S, Park NI, Park SU.** 2021. Enhancement of the flavone contents of *Scutellaria baicalensis* hairy roots via metabolic engineering using maize Lc and Arabidopsis PAP1 transcription factors. *Metabolic Engineering* **64**, 64–73.
- Pasin F, Bedoya LC, Bernabé-Orts JM, Gallo A, Simón-Mateo C, Orzaez D, García JA.** 2017. Multiple T-DNA Delivery to Plants Using Novel Mini Binary Vectors with Compatible Replication Origins. *ACS Synthetic Biology* **6**, 1962–1968.
- Pastor V, Sánchez-Bel P, Gamir J, Pozo MJ, Flors V.** 2018. Accurate and easy method for systemin quantification and examining metabolic changes under different endogenous levels. *Plant Methods* **14**, 33–33.
- Patron NJ.** 2016. Blueprints for green biotech: development and application of standards for plant synthetic biology. *Biochemical Society Transactions* **44**, 702–708.



- Patron NJ.** 2020. Beyond natural: synthetic expansions of botanical form and function. *The New Phytologist* **227**, 295–310.
- Patron NJ, Orzaez D, Marillonnet S, et al.** 2015. Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *The New phytologist* **208**, 13–19.
- Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL.** 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**, 290–295.
- Petolino JF.** 2015. Genome editing in plants via designed zinc finger nucleases. *In vitro cellular & developmental biology. Plant : journal of the Tissue Culture Association* **51**, 1–8.
- Petolino JF, Davies JP.** 2013. Designed transcriptional regulators for trait development. *Plant Science* **201–202**, 128–136.
- Piatek A, Ali Z, Baazim H, Li L, Abulfaraj A, Al-Shareef S, Aouida M, Mahfouz MM.** 2015. RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors. *Plant Biotechnology Journal* **13**, 578–589.
- Platt RJ, Chen S, Zhou Y, et al.** 2014. CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. *Cell* **159**, 440–455.
- Pollak B, Cerda A, Delmans M, Álamos S, Moyano T, West A, Gutiérrez RA, Patron NJ, Federici F, Haseloff J.** 2019. Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytologist* **222**, 628–640.

- Pomerening JR, Kim SY, Ferrell JE.** 2005. Systems-Level Dissection of the Cell-Cycle Oscillator: Bypassing Positive Feedback Produces Damped Oscillations. *Cell* **122**, 565–578.
- Pouvreau B, Vanhercke T, Singh S.** 2018. From plant metabolic engineering to plant synthetic biology: The evolution of the design/build/test/learn cycle. *Synthetic Biology Meets Plant Metabolism* **273**, 3–12.
- Pudasaini A, El-Arab KK, Zoltowski BD.** 2015. LOV-based optogenetic devices: light-driven modules to impart photoregulated control of cellular signaling. *Frontiers in molecular biosciences* **2**, 18–18.
- Purnick PEM, Weiss R.** 2009. The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology* **10**, 410–422.
- Qi H, Blanchard A, Lu T.** 2013*a*. Engineered genetic information processing circuits. *WIREs Systems Biology and Medicine* **5**, 273–287.
- Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA.** 2013*b*. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183.
- Quinlan AR, Hall IM.** 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**, 841–842.
- Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F.** 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**, 2281–2308.

- Randall RS.** 2021. The plant AlcR-pAlcA ethanol-inducible system displays gross growth artefacts independently of downstream pAlcA-regulated inducible constructs. *Scientific Reports* **11**, 2142.
- Rhodijs VA, Segall-Shapiro TH, Sharon BD, et al.** 2013. Design of orthogonal genetic switches based on a crosstalk map of  $\sigma$ s, anti- $\sigma$ s, and promoters. *Molecular Systems Biology* **9**, 702.
- Rizza A, Walia A, Lanquar V, Frommer WB, Jones AM.** 2017. In vivo gibberellin gradients visualized in rapidly elongating tissues. *Nature Plants* **3**, 803–813.
- Robinson MD, McCarthy DJ, Smyth GK.** 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139–140.
- Roca Paixão JF, Gillet F-X, Ribeiro TP, Bournaud C, Lourenço-Tessutti IT, Noriega DD, Melo BP de, de Almeida-Engler J, Grossi-de-Sa MF.** 2019. Improved drought stress tolerance in *Arabidopsis* by CRISPR/dCas9 fusion with a Histone AcetylTransferase. *Scientific Reports* **9**, 8080.
- Roccaro M, Ahmadinejad N, Colby T, Somssich IE.** 2013. Identification of functional cis-regulatory elements by sequential enrichment from a randomized synthetic DNA library. *BMC Plant Biology* **13**, 164.
- Rodriguez PL, Lozano-Juste J.** 2015. Unnatural agrochemical ligands for engineered abscisic acid receptors. *Trends in Plant Science* **20**, 330–332.
- Rollié S, Mangold M, Sundmacher K.** 2012. Designing biological systems: Systems Engineering meets Synthetic Biology. *Chemical Engineering Science* **69**, 1–29.

- Rudge TJ, Brown JR, Federici F, Dalchau N, Phillips A, Ajioka JW, Haseloff J.** 2016. Characterization of Intrinsic Properties of Promoters. *ACS synthetic biology* **5**, 89–98.
- Saijo T, Nagasawa A.** 2014. Development of a tightly regulated and highly responsive copper-inducible gene expression system and its application to control of flowering time. *Plant Cell Reports* **33**, 47–59.
- Samalova M, Brzobohaty B, Moore I.** 2005. pOp6/LhGR: a stringently regulated and highly responsive dexamethasone-inducible gene expression system for tobacco. *The Plant Journal* **41**, 919–935.
- Sanders SM, Ma Z, Hughes JM, et al.** 2018. CRISPR/Cas9-mediated gene knockin in the hydroid *Hydractinia symbiolongicarpus*. *BMC Genomics* **19**, 649.
- Sarrion-Perdigones A, Falconi EE, Zandalinas SI, Juárez P, Fernández-del-Carmen A, Granell A, Orzaez D.** 2011. GoldenBraid: An Iterative Cloning System for Standardized Assembly of Reusable Genetic Modules (J Peccoud, Ed.). *PLoS ONE* **6**, e21622.
- Sarrion-Perdigones A, Vazquez-Vilar M, Palaci J, Castelijn B, Forment J, Ziarsolo P, Blanca J, Granell A, Orzaez D.** 2013. GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *PLANT PHYSIOLOGY* **162**, 1618–1631.
- Schaumberg KA, Antunes MS, Kassaw TK, Xu W, Zalewski CS, Medford JI, Prasad A.** 2016. Quantitative characterization of genetic parts and circuits for plant synthetic biology. *Nature Methods* **13**, 94–100.

- Schoof H, Lenhard M, Haecker A, Mayer KFX, Jürgens G, Laux T.** 2000. The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes. *Cell* **100**, 635–644.
- Schuster M, Kahmann R.** 2019. CRISPR-Cas9 genome editing approaches in filamentous fungi and oomycetes. *Fungal Genetics and Biology* **130**, 43–53.
- Schwechheimer C, Smith C, Bevan MW.** 1998. The activities of acidic and glutamine-rich transcriptional activation domains in plant cells: design of modular transcription factors for high-level expression. *Plant molecular biology* **36**, 195–204.
- Schymanski EL, Gerlich M, Ruttkies C, Neumann S.** 2014. Solving CASMI 2013 with MetFrag, MetFusion and MOLGEN-MS/MS. *Mass spectrometry (Tokyo, Japan)* **3**, S0036–S0036.
- Selma S, Bernabé-Orts JM, Vazquez-Vilar M, Diego-Martin B, Ajenjo M, Garcia-Carpintero V, Granell A, Orzaez D.** 2019. Strong gene activation in plants with genome-wide specificity using a new orthogonal CRISPR/Cas9-based programmable transcriptional activator. *Plant biotechnology journal* **17**, 1703–1705.
- Selma S, Orzáez D.** 2021. Perspectives for epigenetic editing in crops. *Transgenic Research* **30**, 381–400.
- Selma S, Sanmartín N, Espinosa-Ruiz A, Gianoglio S, Lopez-Gresa M, Vázquez-Vilar M, Flors V, Granell A, Orzaez D.** 2021. *Custom-made design of metabolite composition in N. benthamiana leaves using CRISPR activators.* *Synthetic Biology.*

- Senís E, Fatouros C, Große S, Wiedtke E, Niopek D, Mueller A-K, Börner K, Grimm D.** 2014. CRISPR/Cas9-mediated genome engineering: An adeno-associated viral (AAV) vector toolbox. *Biotechnology Journal* **9**, 1402–1412.
- Shah DM, Horsch RB, Klee HJ, et al.** 1986. Engineering Herbicide Tolerance in Transgenic Plants. *Science* **233**, 478–481.
- Shakiba N, Jones RD, Weiss R, Del Vecchio D.** 2021. Context-aware synthetic biology by controller design: Engineering the mammalian cell. *Cell Systems* **12**, 561–592.
- Shanidze N, Lenkeit F, Hartig JS, Funck D.** 2020. A Theophylline-Responsive Riboswitch Regulates Expression of Nuclear-Encoded Genes1 [OPEN]. *Plant Physiology* **182**, 123–135.
- Shao X, Wu S, Dou T, et al.** 2020. Using CRISPR/Cas9 genome editing system to create MaGA20ox2 gene-modified semi-dwarf banana. *Plant Biotechnology Journal* **18**, 17–19.
- Shelton AM, Olmstead DL, Burkness EC, Hutchison WD, Dively G, Welty C, Sparks AN.** 2013. Multi-State Trials of Bt Sweet Corn Varieties for Control of the Corn Earworm (Lepidoptera: Noctuidae). *Journal of Economic Entomology* **106**, 2151–2159.
- Shi T-Q, Gao J, Wang W-J, Wang K-F, Xu G-Q, Huang H, Ji X-J.** 2019. CRISPR/Cas9-Based Genome Editing in the Filamentous Fungus *Fusarium fujikuroi* and Its Application in Strain Engineering for Gibberellic Acid Production. *ACS Synthetic Biology* **8**, 445–454.
- Shrestha A, Khan A, Dey N.** 2018. cis–trans Engineering: Advances and Perspectives on Customized Transcriptional Regulation in Plants. *Molecular Plant* **11**, 886–898.

- Silverstone AL, Jung H-S, Dill A, Kawaide H, Kamiya Y, Sun T.** 2001. Repressing a Repressor: Gibberellin-Induced Rapid Reduction of the RGA Protein in Arabidopsis. *The Plant Cell* **13**, 1555–1566.
- Slusarczyk AL, Lin A, Weiss R.** 2012. Foundations for the design and implementation of synthetic genetic circuits. *Nature Reviews Genetics* **13**, 406–420.
- Small I.** 2007. RNAi for revealing and engineering plant gene functions. *Plant biotechnology / Food biotechnology* **18**, 148–153.
- Smith HO, Hutchison CA, Pfannkoch C, Venter JC.** 2003. Generating a synthetic genome by whole genome assembly:  $\phi$ X174 bacteriophage from synthetic oligonucleotides. *Proceedings of the National Academy of Sciences* **100**, 15440–15445.
- Smolke CD.** 2009. Building outside of the box: iGEM and the BioBricks Foundation. *Nature Biotechnology* **27**, 1099–1102.
- Spelt C, Quattrocchio F, Mol JN, Koes R.** 2000. anthocyanin1 of petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. *The Plant cell* **12**, 1619–1632.
- Spitzer-Rimon B, Farhi M, Albo B, et al.** 2012. The R2R3-MYB-like regulatory factor EOBI, acting downstream of EOBI1, regulates scent production by activating ODO1 and structural scent-related genes in petunia. *The Plant cell* **24**, 5089–5105.
- Stege JT, Guan X, Ho T, Beachy RN, Barbas CF 3rd.** 2002. Controlling gene expression in plants using synthetic zinc finger transcription factors. *The Plant journal : for cell and molecular biology* **32**, 1077–1086.

- Stepanova AN, Yun J, Likhacheva AV, Alonso JM.** 2007. Multilevel Interactions between Ethylene and Auxin in Arabidopsis Roots. *The Plant Cell* **19**, 2169–2185.
- Stephenson MJ, Osbourn A.** 2020. Making drugs out of sunlight and ‘thin air’: an emerging synergy of synthetic biology and natural product chemistry. *The Biochemist* **42**, 34–39.
- Stewart CN, Patron N, Hanson AD, Jez JM.** 2018. Plant metabolic engineering in the synthetic biology era: plant chassis selection. *Plant Cell Reports* **37**, 1357–1358.
- Strelkova N, Barahona M.** 2011. Transient dynamics around unstable periodic orbits in the generalized repressilator model. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **21**, 023104.
- Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD.** 2014*a*. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*.
- Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD.** 2014*b*. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646.
- Tang X, Levi G, Lowder, Zhang T, et al.** 2017. A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nature Plants* **3**, 1–5.
- Tiwari SB, Belachew A, Ma SF, et al.** 2012. The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors: Strong plant transcriptional activation domain. *The Plant Journal* **70**, 855–865.
- Tiwari SB, Hagen G, Guilfoyle TJ.** 2004. Aux/IAA proteins contain a potent transcriptional repression domain. *The Plant cell* **16**, 533–543.



- Torti S, Schlesier R, Thümmeler A, et al.** 2021. Transient reprogramming of crop plants for agronomic performance. *Nature Plants* **7**, 159–171.
- Tuncel A, Corbin KR, Ahn-Jarvis J, Harris S, Hawkins E, Smedley MA, Harwood W, Warren FJ, Patron NJ, Smith AM.** 2019. Cas9-mediated mutagenesis of potato starch-branching enzymes generates a range of tuber starch phenotypes. *Plant Biotechnology Journal* **17**, 2259–2271.
- Tzfira T, Citovsky V.** 2006. Agrobacterium-mediated genetic transformation of plants: biology and biotechnology. *Plant biotechnology/Food biotechnology* **17**, 147–154.
- Unger E, Cigan AM, Trimnell M, Xu R, Kendall T, Roth B, Albertsen M.** 2002. A Chimeric Ecdysone Receptor Facilitates Methoxyfenozide-Dependent Restoration of Male Fertility in ms45 Maize. *Transgenic Research* **11**, 455–465.
- Uranga M, Aragonés V, Selma S, Vázquez-Vilar M, Orzáez D, Daròs J-A.** 2020. Efficient Cas9 multiplex editing using unspaced gRNA arrays engineering in a Potato virus X vector. *bioRxiv*, 2020.06.25.170977.
- Uranga M, Aragonés V, Selma S, Vázquez-Vilar M, Orzáez D, Daròs J-A.** 2021a. Efficient Cas9 multiplex editing using unspaced sgRNA arrays engineering in a Potato virus X vector. *The Plant Journal: For Cell and Molecular Biology* **106**, 555–565.
- Uranga M, Vazquez-Vilar M, Orzáez D, Daròs J-A.** 2021b. CRISPR-Cas12a Genome Editing at the Whole-Plant Level Using Two Compatible RNA Virus Vectors. *The CRISPR Journal* **4**, 761–769.
- Vavitsas K, Crozet P, Vinde MH, Davies F, Lemaire SD, Vickers CE.** 2019. The Synthetic Biology Toolkit for Photosynthetic Microorganisms. *Plant physiology* **181**, 14–27.

- Vazquez-Vilar M, Bernabé-Orts JM, Fernandez-Del-Carmen A, Ziarsolo P, Blanca J, Granell A, Orzaez D.** 2016. A modular toolbox for gRNA-Cas9 genome engineering in plants based on the GoldenBraid standard. *Plant methods* **12**, 10.
- Vazquez-Vilar M, Gandía M, García-Carpintero V, Marqués E, Sarrion-Perdigones A, Yenush L, Polaina J, Manzanares P, Marcos JF, Orzaez D.** 2020. Multigene Engineering by GoldenBraid Cloning: From Plants to Filamentous Fungi and Beyond. *Current Protocols in Molecular Biology*.
- Vazquez-vilar M, Quijano-rubio A, Fernandez-del-carmen A, Sarrion-perdigones A, Ochoa-fernandez R, Ziarsolo P, Granell A, Orzaez D.** 2017. GB3 . 0 : a platform for plant bio-design that connects functional DNA elements with associated biological data. *45*, 2196–2209.
- Veillet F, Perrot L, Chauvin L, Kermarrec M-P, Guyon-Debast A, Chauvin J-E, Nogué F, Mazier M.** 2019. Transgene-Free Genome Editing in Tomato and Potato Plants Using Agrobacterium-Mediated Delivery of a CRISPR/Cas9 Cytidine Base Editor. *International Journal of Molecular Sciences* **20**.
- Venter M.** 2007. Synthetic promoters: genetic control through cis engineering. *Trends in plant science* **12**, 118–124.
- Verdonk JC, Haring MA, van Tunen AJ, Schuurink RC.** 2005. ODORANT1 regulates fragrance biosynthesis in petunia flowers. *The Plant cell* **17**, 1612–1624.
- Vu AT, Lee JM.** 2019. Genetic variations underlying anthocyanin accumulation in tomato fruits. *Euphytica* **215**, 196.

- Waadt R, Hitomi K, Nishimura N, Hitomi C, Adams SR, Getzoff ED, Schroeder JI.** 2014. FRET-based reporters for the direct visualization of abscisic acid concentration changes and distribution in Arabidopsis. *eLife* **3**, e01739–e01739.
- Wang X.** 2009. Structure, mechanism and engineering of plant natural product glycosyltransferases. *FEBS Letters* **583**, 3303–3309.
- Wang B, Kitney RI, Joly N, Buck M.** 2011. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications* **2**, 508.
- Waryah CB, Moses C, Arooj M, Blancafort P.** 2018. Zinc Fingers, TALEs, and CRISPR Systems: A Comparison of Tools for Epigenome Editing. In: Jeltsch A,, In: Rots MG, eds. *Epigenome Editing: Methods and Protocols*. New York, NY: Springer New York, 19–63.
- Way JC, Collins JJ, Keasling JD, Silver PA.** 2014. Integrating Biological Redesign: Where Synthetic Biology Came From and Where It Needs to Go. *Cell* **157**, 151–161.
- Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S.** 2011. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE* **6**, e16765.
- Wegmann U, Carvalho AL, Stocks M, Carding SR.** 2017. Use of genetically modified bacteria for drug delivery in humans: Revisiting the safety aspect. *Scientific Reports* **7**, 2294.
- Weinmann P, Gossen M, Hillen W, Bujard H, Gatz C.** 1994. A chimeric transactivator allows tetracycline-responsive gene expression in whole plants. *The Plant Journal* **5**, 559–569.

- Wend S, Bosco CD, Kämpf MM, Ren F, Palme K, Weber W, Dovzhenko A, Zurbriggen MD.** 2013. A quantitative ratiometric sensor for time-resolved analysis of auxin dynamics. *Scientific Reports* **3**, 2052.
- Wu S, Zhu H, Liu J, Yang Q, Shao X, Bi F, Hu C, Huo H, Chen K, Yi G.** 2020. Establishment of a PEG-mediated protoplast transformation system based on DNA and CRISPR/Cas9 ribonucleoprotein complexes for banana. *BMC Plant Biology* **20**, 425.
- Wydro M, Kozubek E, Lehmann P.** 2006. Optimization of transient *Agrobacterium*-mediated gene expression system in leaves of *Nicotiana benthamiana*. *Acta biochimica Polonica* **53**, 289–298.
- Xie Q, Liu Z, Meir S, Rogachev I, Aharoni A, Klee HJ, Galili G.** 2016. Altered metabolite accumulation in tomato fruits by coexpressing a feedback-insensitive AroG and the PhODO1 MYB-type transcription factor. *Plant biotechnology journal* **14**, 2300–2309.
- Xie D-Y, Sharma SB, Wright E, Wang Z-Y, Dixon RA.** 2006. Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor. *The Plant Journal* **45**, 895–907.
- Xing H-L, Dong L, Wang Z-P, Zhang H-Y, Han C-Y, Liu B, Wang X-C, Chen Q-J.** 2014. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology* **14**, 327.
- Xiong X, Liang J, Li Z, Gong B-Q, Li J-F.** 2021. Multiplex and optimization of dCas9-TV-mediated gene activation in plants. *Journal of Integrative Plant Biology* **63**, 634–645.
- Xu CL, Ruan MZC, Mahajan VB, Tsang SH.** 2019. Viral Delivery Systems for CRISPR. *Viruses* **11**.

- Yin K, Han T, Liu G, Chen T, Wang Y, Yu AYL, Liu Y.** 2015. A geminivirus-based guide RNA delivery system for CRISPR/Cas9 mediated plant genome editing. *Scientific reports* **5**, 14926–14926.
- You Y-S, Marella H, Zentella R, Zhou Y, Ulmasov T, Ho T-HD, Quatrano RS.** 2006. Use of Bacterial Quorum-Sensing Components to Regulate Gene Expression in Plants. *Plant Physiology* **140**, 1205–1212.
- Zalatan JG, Lee ME, Almeida R, et al.** 2015. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350.
- Zaynab M, Sharif Y, Fatima M, et al.** 2020. CRISPR/Cas9 to generate plant immunity against pathogen. *Microbial Pathogenesis* **141**, 103996.
- Zhang JZ.** 2003. Overexpression analysis of plant transcription factors. *Current Opinion in Plant Biology* **6**, 430–440.
- Zhang Y, Butelli E, Alseekh S, et al.** 2015*a*. Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nature Communications* **6**, 1–11.
- Zhang X, He Y, Li L, Liu H, Hong G.** 2021*a*. Involvement of the R2R3-MYB transcription factor MYB21 and its homologs in regulating flavonol accumulation in Arabidopsis stamen. *Journal of experimental botany* **72**, 4319–4332.
- Zhang Y, Iaffaldano B, Qi Y.** 2021*b*. CRISPR ribonucleoprotein-mediated genetic engineering in plants. *Technology and Applications in Plants* **2**, 100168.

- Zhang Y, Ren Q, Tang X, et al.** 2021c. Expanding the scope of plant genome engineering with Cas12a orthologs and highly multiplexable editing systems. *Nature Communications* **12**, 1944.
- Zhang L, Takahashi Y, Hsu P-K, Kollist H, Merilo E, Krysan PJ, Schroeder JI.** 2020. FRET kinase sensor development reveals SnRK2/OST1 activation by ABA but not by MeJA and high CO<sub>2</sub> during stomatal closure (DC Bergmann, CS Hardtke, and J Leung, Eds.). *eLife* **9**, e56351.
- Zhang X-H, Tee LY, Wang X-G, Huang Q-S, Yang S-H.** 2015b. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids* **4**, e264.
- Zhang Q, Xing H-L, Wang Z-P, Zhang H-Y, Yang F, Wang X-C, Chen Q-J.** 2018. Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in *Arabidopsis* and its prevention. *Plant Molecular Biology* **96**, 445–456.
- Zhang Y, Yin C, Zhang T, et al.** 2015c. CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs. *Scientific Reports* **5**, 16277.
- Zhao L, Gao L, Wang H, Chen X, Wang Y, Yang H, Wei C, Wan X, Xia T.** 2013. The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis. *Functional & Integrative Genomics* **13**, 75–98.
- Zhu X, Liu X, Liu T, Wang Y, Ahmed N, Li Z, Jiang H.** 2021. Synthetic biology of plant natural products: From pathway elucidation to engineered biosynthesis in plant cells. *Plant communications* **2**, 100229–100229.

**Zurbruggen MD, Moor A, Weber W.** 2012. Plant and bacterial systems biology as a platform for plant synthetic bio(techno)logy. *Genome-based Microbiology – From -omics Research to Systems and Synthetic Biology* **160**, 80–90.

**Zürcher E, Tavor-Deslex D, Lituiev D, Enkerli K, Tarr PT, Müller B.** 2013. A Robust and Sensitive Synthetic Sensor to Monitor the Transcriptional Output of the Cytokinin Signaling Network in Planta. *Plant Physiology* **161**, 1066–1075.









