





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

Sara Selma¹ , Neus Sanmartín² , Ana Espinosa-Ruiz¹, Silvia Gianoglio¹, Maria Pilar Lopez-Gresa¹, Marta Vázquez-Vilar¹ , Victor Flors², Antonio Granell¹ and Diego Orzaez^{1,*} 

¹Instituto Biología Molecular de Plantas, CSIC-UPV, Valencia, Spain

²Escuela Superior de Tecnología y Ciencias Experimentales, Universidad Jaume I, Castellón de la Plana, Spain

Received 11 January 2022;

revised 7 April 2022;

accepted 28 April 2022.

*Correspondence (Tel +34 963879933; fax +34 963877879; email dorzaez@ibmcp.upv.es)

Summary

Transcriptional regulators based on CRISPR architecture expand our ability to reprogramme 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 programmes for each of the four metabolites, each programme 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 programme 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: CRISPRa, 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 by 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, and oscillators) (Bernabé-Orts *et al.*, 2020; McKenzie *et al.*, 1998; Ochoa-Fernandez *et al.*, 2016). 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 (Naing *et al.*, 2017; Xie *et al.*, 2006), 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 in CRISPR/Cas has emerged that allow 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 activation domains (ADs) 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 in dCas9 have been described (Li *et al.*, 2019; Lowder *et al.*, 2018; Pan *et al.*, 2021; Park *et al.*, 2017), showing efficient activation of target endogenous genes. The two main advantages of dCas9-PTA are (i) their high accuracy, reaching single-gene specificity levels as shown recently in a transcriptomic analysis showing negligible off-target activation (Li *et al.*, 2017); and (ii) their amenability for multiplexing (Lowder *et al.*, 2017; Vázquez-Vilar *et al.*, 2016). 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 (see Figure 1). The first enzymatic reactions of the pathway generates cinnamic acid, coumaric acid, and 4-coumaroyl-CoA, the basic backbone products derive from phenylalanine, thus providing the core structures for the biosynthesis of all flavonoids as well as for other phenylpropanoid branches like the lignin pathway; 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 will be the substrate for multiple reactions that originate the distinct subgroups of the flavonoid pathway, including flavonols, anthocyanins, and isoflavonoids (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 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.*, 2015). 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 to 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-scrRNA strategy (Koneremann *et al.*, 2015; 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 programme. In the constant module, the first TU produces a deactivated Cas9 translationally fused to an EDLL domain, a plant activation domain isolated from 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 programmes (i.e., single polycistronic constructs expressing up to six gRNAs) for 10 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 programmes (i.e., combinations of polycistronic gRNA constructs targeting several genes simultaneously) were constructed and assayed transiently in *N. benthamiana*, each programme 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.

Results

Optimization of activation programmes 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 (<https://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

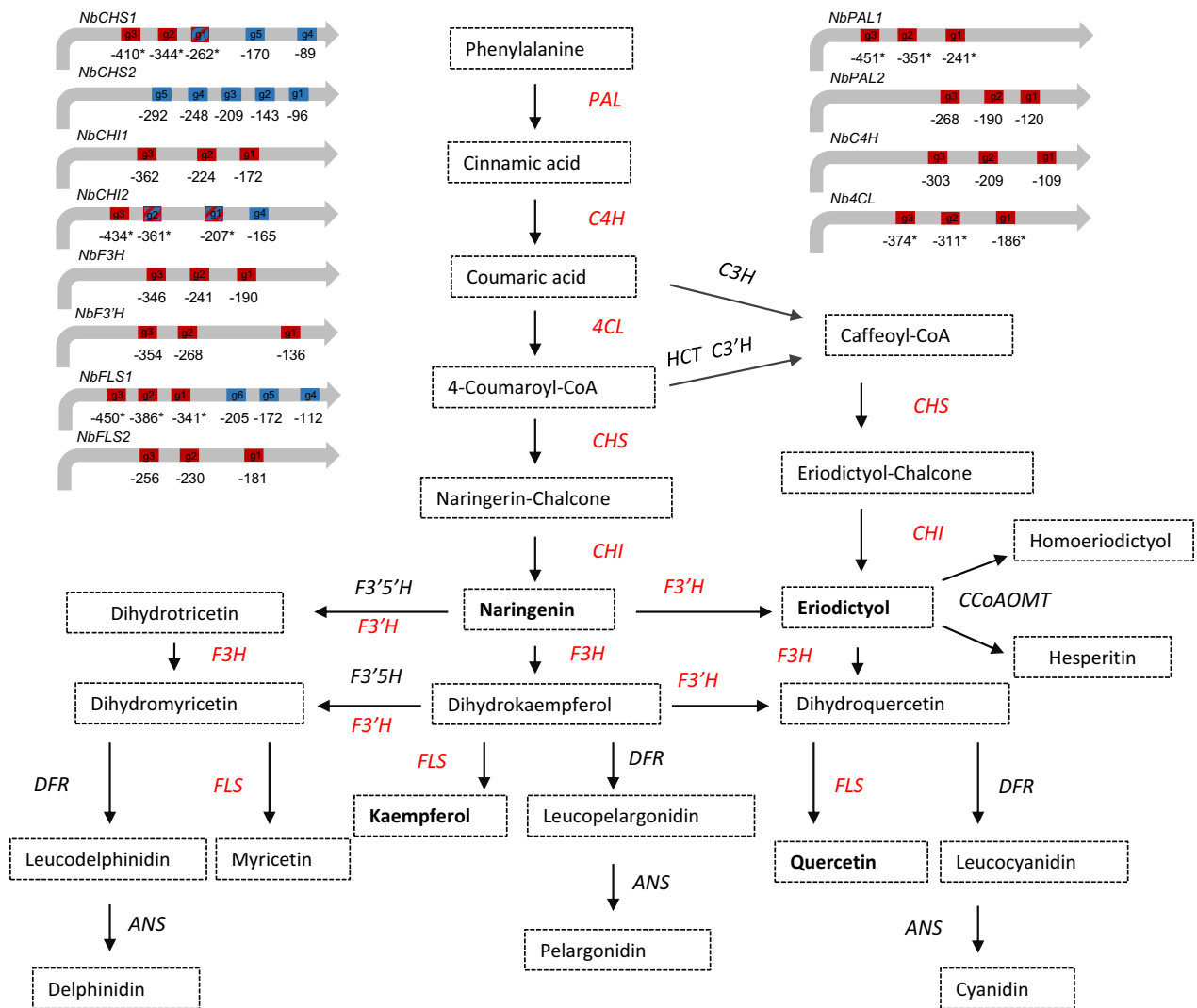


Figure 1 Schematic representation of the flavonoid biosynthetic pathway in plants. The different metabolites are represented in boxes. The genes that integrate the pathway are: ANS (anthocyanidin synthase), CHS (Chalcone synthase), CHI (Chalcone isomerase), CCoAOMT (Caffeoyl-CoA O-methyltransferase), C3H (4-coumarate 3-hydroxylase), C3'H (4-coumaroyl shikimate/quinic 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'-5H (Flavonoid 3' 5'-hydroxylase), F3'-5'H (Flavonoid 3' 5'-hydroxylase), FLS (Flavonol synthase), HCT (4-hydroxycinnamoyl CoA: shikimate/quinic 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 programmes assayed in both optimization rounds. The asterisks represent the gRNA position recalculated with updated information on the TSS available in databases (<https://www.nbent.com/>).

N. benthamiana genome was carried out manually by homology search using orthologous proteins from *A. thaliana*, *Solanum lycopersicum*, and *Nicotiana tabacum* available in Uniprot (<https://www.uniprot.org/>) and Solgenomics database (<https://solgenomics.net/>). The retrieved candidates were contrasted with the automatic annotation of version v3.3 of the *N. benthamiana* genome assembly (<https://www.nbent.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 the transcriptional start site (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, 12 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.

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 a MS2-binding RNA 2.1 aptamer separated by processable tRNA spacers (Figure 2b). All protospacers were designed against the 5'untranscribed region of genes between -100 and -350 bp from the TSS. The list of the gRNAs target sites designed for each gene is listed in Table S1 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 10 selected genes, namely, *NbPAL1*, *NbPAL2*, *NbC4H*, *NbCL4*, *NbCHS1*, *NbCHI2*, *NbF3'H*, *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 Figure S1A). Surprisingly, the activation programme designed for *NbPAL1* resulted in a modest four-fold upregulation, while its homologue *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 (Figure S1A).

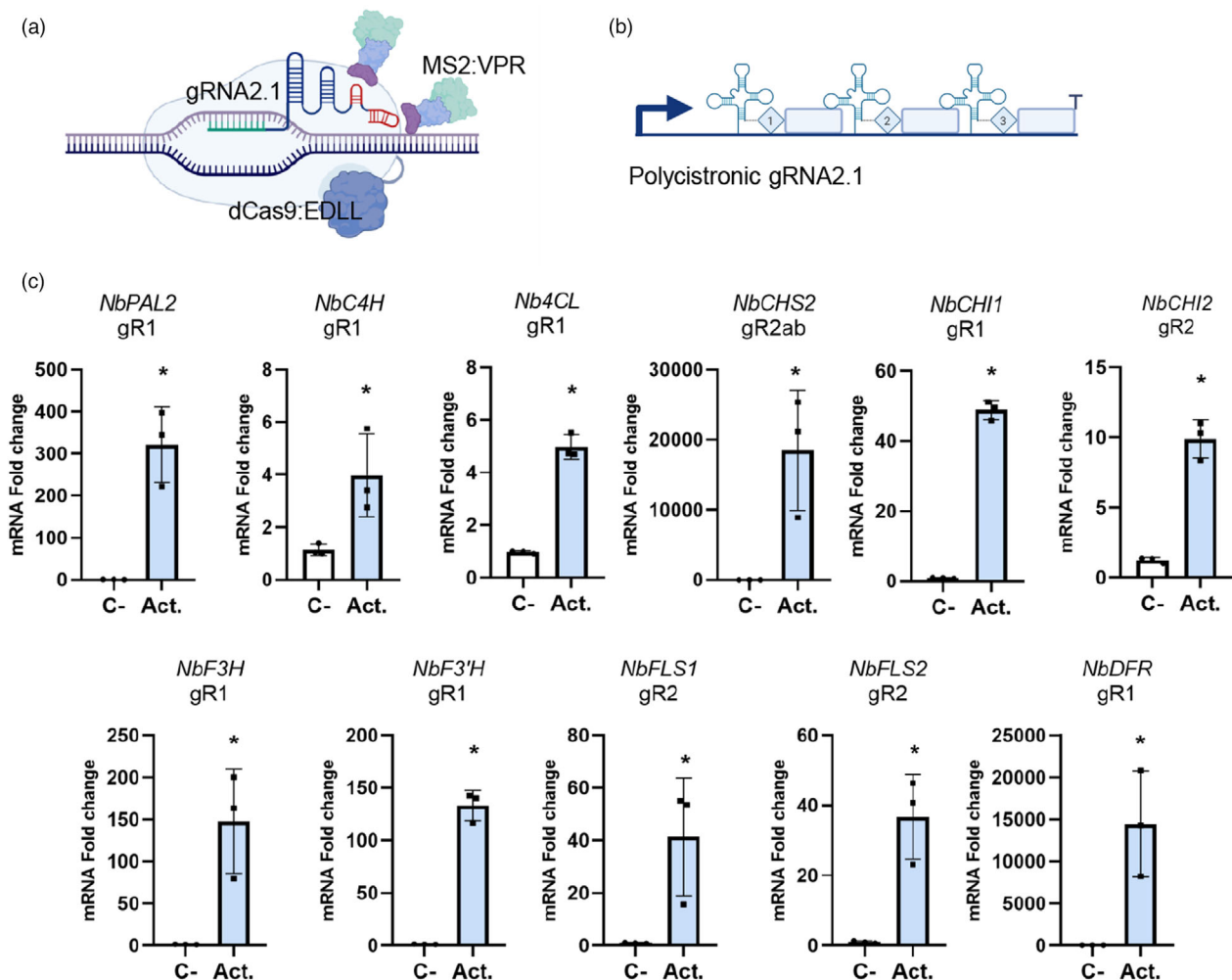


Figure 2 dCasEV2.1-mediated transcriptional activation of individual genes of the flavonoid pathway in *N. benthamiana*. (a) Schematic representation of the dCasEV2.1 activator, comprising the dCas9 fused to the EDLL activation domain in C-terminus and the coat protein of the phage MS2 fused to the activation 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 indicates 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.

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 Figure S1B. 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 programmes to be employed in subsequent experiments. The final set of activation programmes for each gene was then re-evaluated in a confirmatory experiment represented in Figure 2c. As can be observed, successful transcriptional activations (>10-fold) were obtained for all selected enzymatic steps except for *NbC4H* and *Nb4CL*, where only modest four-fold activation rates were achieved. The best activation results were achieved for the previously described *NbDFR* (15 000-fold) and *NbCHS2* (18 000-fold).

Programming naringenin accumulation with dCasEV2.1

The results obtained with the individual activation programmes prompted us to undertake the simultaneous activation of several genes in the pathway following a modular polycistronic gRNA strategy (Hashimoto *et al.*, 2018; Lowder *et al.*, 2017). 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 programme 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 programmes 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 programmes resulted in significant gene activation, although induction levels were notably reduced compared with programmes 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 programmes AP-N1, AP-N1B, AP-N2, and AP-N3 comprising 2 or 3 target genes but dropped below 600-fold in programmes targeting four or five genes (AP-N2B and AP-N4B). Following the same trend, those targets showing modest activations in single-gene programmes showed even lower inductions with complex activation programmes, 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 programme. 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 in 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 programmes (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 in the later time point, particularly in samples AP-N3, AP-N4, and AP-N4B.

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

Despite the indications that complex gRNA programmes resulted in lower activation efficiencies than simpler ones, the remarkable levels of naringenin accumulation obtained with shorter programmes 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 programme as a basis, and adding activation instructions for *F3H* and *FLS*, producing first dihydroquercetin and finally quercetin. Following this rationale, three new metabolite-specific gRNA programmes (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 gene activation profiles obtained for each combination, where it can be observed that an increase in programme complexity has a negative effect on the overall induction levels compared to simpler programmes. This was clearly observed in the activation levels of *NbCHS2*, which drops by approximately five-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 programme (AP-K and AP-Q). It is worth noticing that *NbF3'H* is only activated in AP-Q but not in AP-K, this serving as an additional indication of the specificity of the activation programmes and discarding positive feedback as 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 programme 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 Programmes to the flavonols Activation Programmes accounting

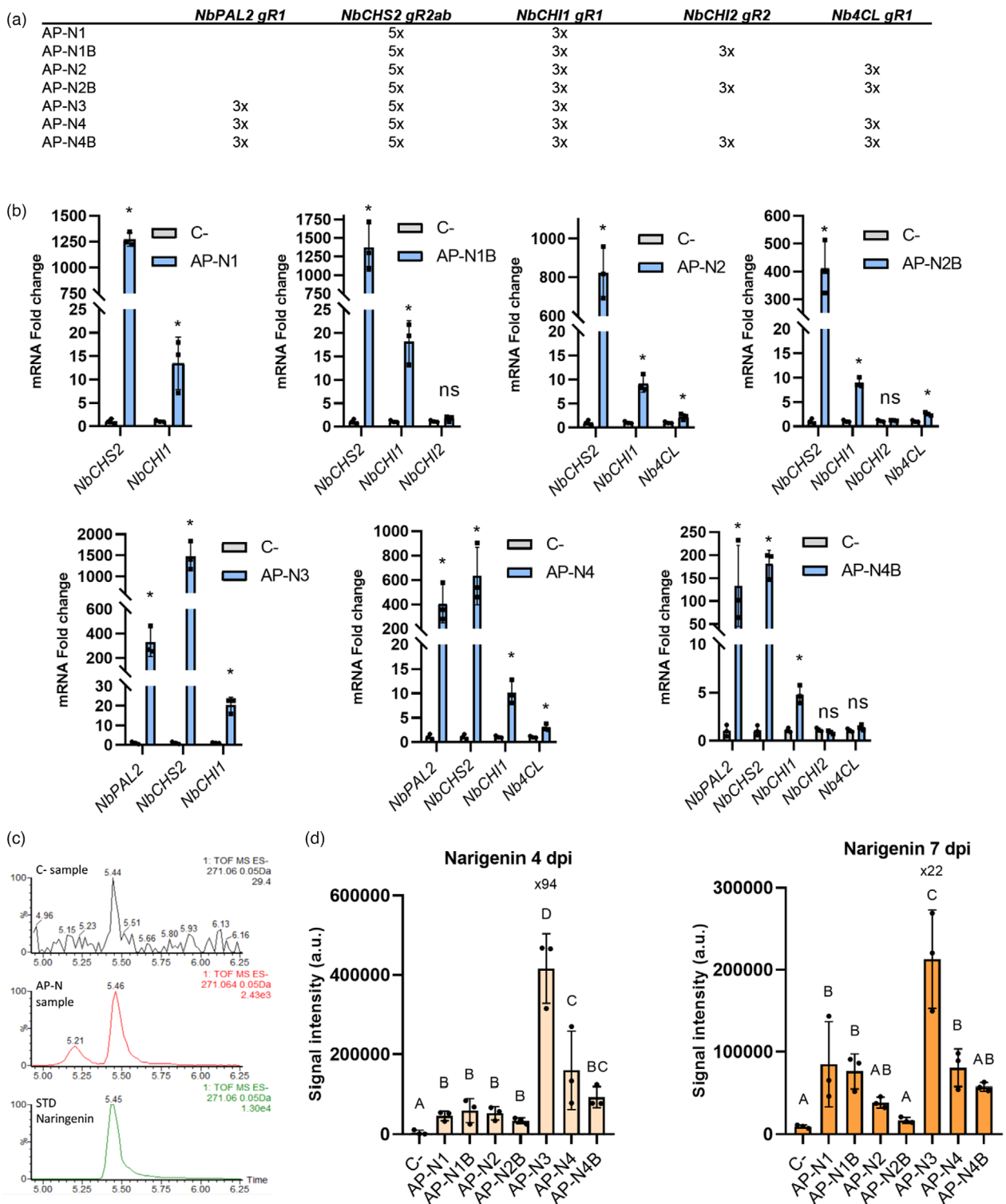


Figure 3 Optimization of naringenin production in *N. benthamiana* leaves through dCasEV2.1 activation. (a) Representation of the different activation programmes (APs) tested for activating naringenin production (AP-N1, AP-N1B, AP-N2, AP-N2B, AP-N3, AP-N4, and AP-N4B). In the table, Nx values represent 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 ($* = P < 0.05$).

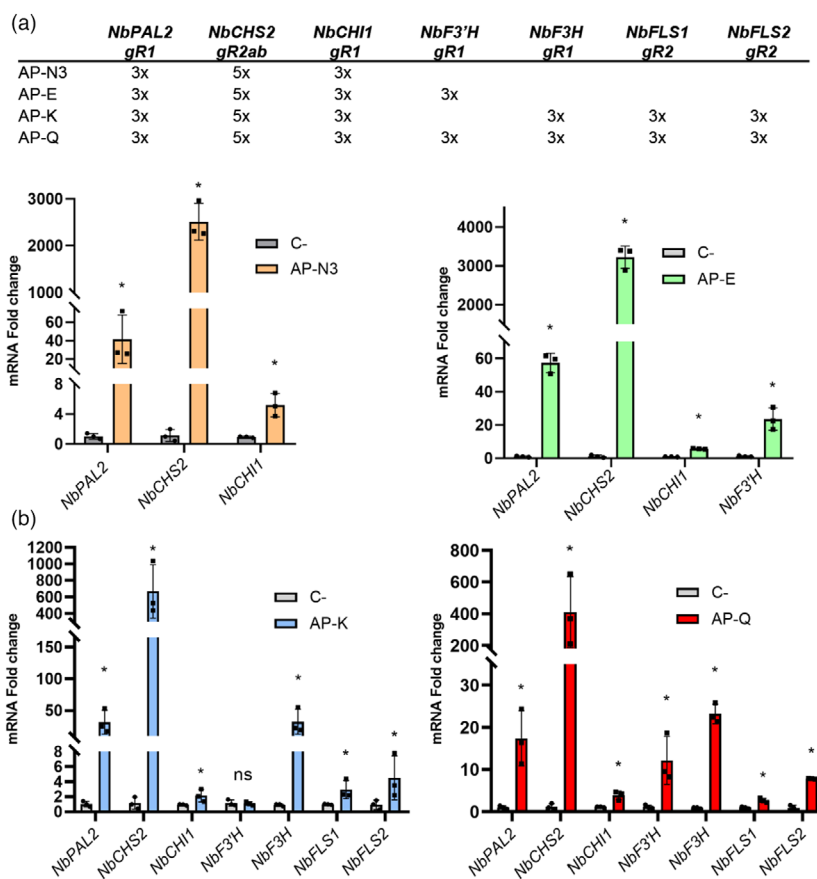


Figure 4 Design and relative expression analyses of activation programmes (APs) for naringenin, eriodictyol, kaempferol, and quercetin production in *N. benthamiana*. (a) Representation of the Activation Programmes (APs), the gRNAs and the targeted genes included in each of them (AP-N3 for Naringenin Activation Programme, AP-E for Eriodictyol Activation Programme, AP-K for Kaempferol Activation Programme, and AP-Q for Quercetin Activation Programme). 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 ($* = P < 0.05$).

for 23.4% of the variance. The level of precision achieved with the four activation programmes 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 programmes and the control sample. A detailed version of the heat map can be found in Figure S3. The differential m/z ions and their respective retention times are listed in Table S2. 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 programme 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 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 programmes. Again, this is not entirely unexpected as both AP-K and AP-Q share 5 out of 6 steps in their respective programmes. 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 a first generation of PTAs based on the direct translational fusion of ADs 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 (Lee *et al.*, 2019; Lowder *et al.*, 2018). These improvements, added to the multiplexing capacity of CRISPR/dCas9, have turned CRISPR-PTAs into extraordinary tools for Plant Synthetic Biology applications. Of particular

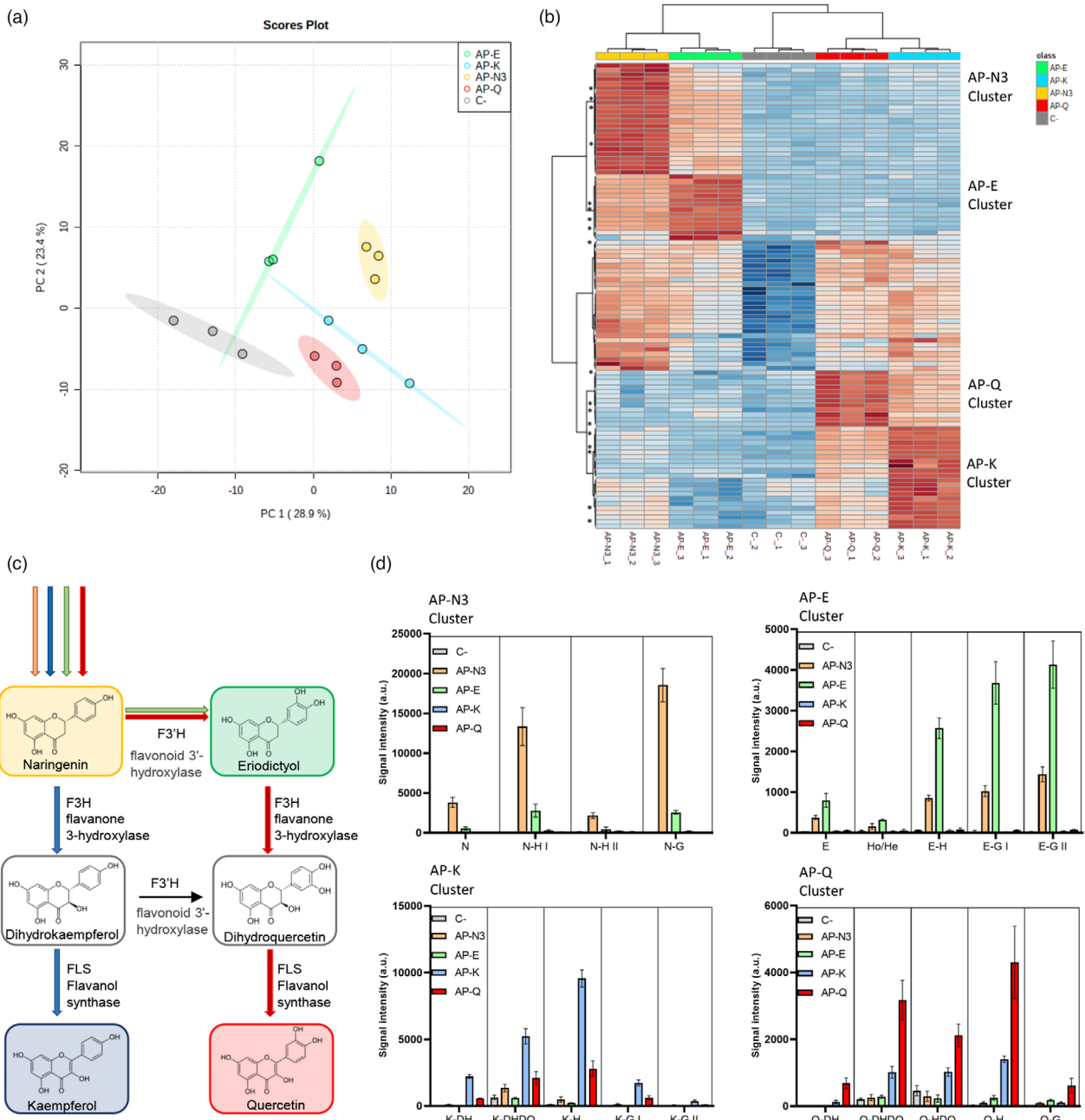


Figure 5 Analysis of the metabolic profiles of *N. benthamiana* leaves treated with activation programmes (APs) for naringenin, eriodictyol, kaempferol, and quercetin. (a) Principal Component Analysis resulting from the untargeted LC-MS data obtained in ESI negative mode from leaves treated with programmes AP-N3, AP-E, AP-K, and AP-Q and control samples (C-) agroinfiltrated with unprogrammed dCasEV2.1. (b) Hierarchical cluster analysis and heat map 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 significantly different using an ANOVA analysis ($P < 0.05$). The data were 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 heat map as single compounds (See also Table S2). 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.

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.*, 2021). 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 to refine the accuracy of re-routing programmes. 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 programme transcriptional activation, its combination with efficient repressors would enable near full control of metabolic pathways in plants.

For the delivery of activation programmes, 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 (Grosse-Holz *et al.*, 2018; Norkunas *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 programmes 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 programmes 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 programmes with endogenous or exogenous cues using appropriated 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 sum, we show here that CRISPR/dCas9-based transcriptional activators provide sufficient precision and multiplexing capacity to reprogramme metabolic pathways and customize metabolic composition in plants. This ability has important implications in feed and food design, as well as in the valorization of industrial crops.

Experimental procedures

gRNA design

The selection of the protospacer sequences for the gRNA design was performed following the criteria previously described in 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 (<https://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 Table S1.

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 <https://gbcloning.upv.es/> by entering the IDs provided in Table S3. 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 https://gbcloning.upv.es/do/crispr/multi_cas9_gRNA_domesticator_1. Primers were resuspended in water to final concentrations of 1 μM 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. One microlitre 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. The next step consists of the assembly of the multiplexing gRNA expression cassette in level 1. For level 1 assemblies, 75 ng of level 0 gRNA for each position, 75 ng of U626 promoter (GB1001), and 75 ng of pDGB3 α destination vector were included in a BsaI restriction-ligation reaction.

The combination of level 1 multiplexing gRNAs and dCasEV2.1 TUs were performed by binary BsaI or BsmBI restriction-ligation reactions to obtain all the level ≥ 1 assemblies as it was 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 growing chamber where the growing conditions were 24 °C/20 °C light/darkness with a 16 h/8 h photoperiod. The plasmids were transferred to *A. tumefaciens* strain GV3101 by electroporation. Agroinfiltration was carried out with overnight grown bacterial cultures. The cultures were pelleted and resuspended in agroinfiltration solution (10 mM MES, pH 5.6, 10 mM MgCl₂, 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 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 agroinfiltrated leaves using the Gene Jet Plant Purification Mini Kit (Thermo Fisher Scientific, Massachusetts, USA) 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, and 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, Kusatsu, Japan) 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 (TB Green® Premix Ex Taq) using Applied biosystem 7500 Fast Real-Time PCR system with specific primer pairs (Table S4). The

qPCR conditions employed are those recommended by the manufacturer (TB Green® Premix Ex Taq -Takara): Hold stage 1 cycle 95 °C 30 s, Cycling Stage 40 Cycles 95 °C 3 s and 60 °C 30 s. 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 the Table S4. 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 an unspecific gRNA. mRNA fold change calculations for each sample were carried out according to the comparative $\Delta\Delta\text{C}_T$ 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, respectively and used as triplicates for metabolomics analyses. The same tissue was used for transcriptomics and metabolomics analyses. The tissue was frozen in liquid N₂ and powdered with a grinding mill and, finally, lyophilized. Thirty mg of dried tissue were extracted at 4 °C with 1 mL of methanol 30% containing 0.01% of formic acid. The preparation was homogenized with a grinding mill and kept on ice for 20 min. After that, the samples were centrifugated at 15 000 rpm for 15 min. The supernatant was collected and filtrated with 20 microns cellulose strainer (Regenerated Cellulose Filter; Teknokroma, Barcelona, Spain). Three independent biological and two technical replicates per sample were analysed. 20 μL of each sample were injected into an Acquity UPLC system (Waters, Massachusetts, USA) 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 \times 100 mm (Waters, Massachusetts, USA). 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 min. 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 a 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, Kawasaki, Japan) 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 (Figure S2).

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 (<https://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 and 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 (<https://www.metaboanalyst.ca/>). Logarithmic transformation and Pareto scaling were employed as normalization to elaborate the Principal Component Analysis and the Hierarchical Clustering and Heat map. Euclidean distance and Ward Clustering algorithm were applied as parameters for elaborating Hierarchical Clustering, and an ANOVA test was the statistical method used for the 100 m/z more significantly different.

The m/z values obtained as significantly different were manually clustered into single metabolites employing the original chromatograms. Finally, the tentative identification of each metabolite was carried out using external databases (<https://pubchem.ncbi.nlm.nih.gov/>) and the information obtained with their fragmentation profiles and collected in Table S2. The quantification of the metabolites was performed employing the parental ion identified in the F1.

Author contributions

D.O. and S.S. designed the experiments. S.S. and N.S. conducted the experiments. A-R.A. and L-G.M. contributes to data analysis. D.O. and S.S. drafted the manuscript. G.S. V-V.M. F.V., and G.T. contribute to the manuscript writing and editing. All the authors discussed and revised the manuscript.

Acknowledgements

This work has been funded by Grant PID2019-108203RB-10 Plan Nacional I + D, Spanish Ministry of Science and Innovation and Spanish Ministry of Economy and Competitiveness. Sara Selma is a recipient of FPI fellowship associated with this Grant (BIO2016-78601-R).

Conflict of interest

The authors declare no conflicts of interest.

Data Availability Statement

The datasets and chromatograms generated in this work are located: <https://doi.org/10.5281/zenodo.5091654>.

References

- Bernabé-Orts, J.M., Quijano-Rubio, A., Vazquez-Vilar, M., Mancheño-Bonillo, J., Moles-Casas, V., Selma, S., Gianoglio, S. et al. (2020) A memory switch for plant synthetic biology based on the phage ϕ C31 integration system. *Nucleic Acids Res.* **48**, 3379–3394.
- Butelli, E., Titta, L., Giorgio, M., Mock, H.-P., Matros, A., Peterek, S., Schijlen, E.G.W.M. et al. (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* **26**, 1301–1308.
- Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S. et al. (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods*, **12**, 326–328.
- Chen, J., Clinton, M., Qi, G., Wang, D., Liu, F. and Fu, Z.Q. (2020) Reprogramming and remodeling: transcriptional and epigenetic regulation of salicylic acid-mediated plant defense. *J. Exp. Bot.* **71**, 5256–5268.
- D'Aoust, M.-A., Lavoie, P.-O., Couture, M.-M.-J., Trépanier, S., Guay, J.-M., Dargis, M., Mongrand, S. et al. (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 Biotechnol. J.* **6**, 930–940.
- Dias, A.P. and Grotewold, E. (2003) Manipulating the accumulation of phenolics in maize cultured cells using transcription factors. *Biochem. Eng. J.* **14**, 207–216.
- Diego-Martin, B., González, B., Vazquez-Vilar, M., Selma, S., Mateos-Fernández, R., Gianoglio, S., Fernández-del-Carmen, A. et al. (2020) Pilot production of SARS-CoV-2 related proteins in plants: a proof of concept for rapid repurposing of indoor farms into biomanufacturing facilities. *Front. Plant Sci.* **11**, 2101.
- Fresquet-Corrales, S., Roque, E., Sarrion-Perdigones, A., Rochina, M., López-Gresa, M.P., Díaz-Mula, H.M., Bellés, J.M. et al. (2017) Metabolic engineering to simultaneously activate anthocyanin and proanthocyanidin biosynthetic pathways in *Nicotiana* spp. *PLoS One*, **12**, e0184839.
- Gallego-Bartolomé, J. (2020) DNA methylation in plants: mechanisms and tools for targeted manipulation. *New Phytol.* **227**, 38–44.
- Gamir, J., Pastor, V., Cerezo, M. and Flors, V. (2012) Identification of indole-3-carboxylic acid as mediator of priming against *Plectosphaerella cucumerina*. *Plant Physiol. Biochem.* **61**, 169–179.
- Ghoshal, B., Yong, B., Picard, C.L., Feng, S., Tam, J.M. and Jacobsen, S.E. (2020) A viral guide RNA delivery system for CRISPR-based transcriptional activation and heritable targeted DNA demethylation in *Arabidopsis thaliana*. *PLoS Genet.* **16**, e1008983.
- Godwin, I.D., Rutkoski, J., Varshney, R.K. and Hickey, L.T. (2019) Technological perspectives for plant breeding. *Theor. Appl. Genet.* **132**, 555–557.
- Grosse-Holz, F., Madeira, L., Zahid, M.A., Songer, M., Kourelis, J., Fesenko, M., Ninck, S. et al. (2018) Three unrelated protease inhibitors enhance accumulation of pharmaceutical recombinant proteins in *Nicotiana benthamiana*. *Plant Biotechnol. J.* **16**, 1797–1810.
- Hashimoto, R., Ueta, R., Abe, C., Osakabe, Y. and Osakabe, K. (2018) Efficient multiplex genome editing induces precise, and self-ligated type mutations in tomato plants. *Front. Plant Sci.* **9**, 916.
- Jisha, V., Dampanaboina, L., Vadassery, J., Mithöfer, A., Kappara, S. and Ramanan, R. (2015) Overexpression of an AP2/ERF type transcription factor OsEREBP1 confers biotic and abiotic stress tolerance in rice. *PLoS One*, **10**, e0127831.
- Khan, M.K., Zill-E-Huma, and Dangles, O. (2014) A comprehensive review on flavanones, the major citrus polyphenols. *J. Food Compos. Anal.* **33**, 85–104.
- Koneremann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D. et al. (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, **517**, 583–588.
- Lee, J.E., Neumann, M., Duro, D.I. and Schmid, M. (2019) CRISPR-based tools for targeted transcriptional and epigenetic regulation in plants. *PLoS One*, **14**, e0222778.
- Li, Z., Wang, F. and Li, J.-F. (2019) Targeted transcriptional activation in plants using a potent dead Cas9-derived synthetic gene activator. *Curr. Protoc. Mol. Biol.* **127**, e89.

- Li, Z., Zhang, D., Xiong, X., Yan, B., Xie, W., Sheen, J. and Li, J.-F. (2017) A potent Cas9-derived gene activator for plant and mammalian cells. *Nat. Plants*, **3**, 930–936.
- Liu, D., Shi, L., Han, C., Yu, J., Li, D. and 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, J., Osbourn, A. and Ma, P. (2015) MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant*, **8**, 689–708.
- Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCT} method. *Methods*, **25**, 402–408.
- Llorente, B., Torres-Montilla, S., Morelli, L., Florez-Sarasa, I., Matus, J.T., Ezquerro, M., D'Andrea, L. et al. (2020) Synthetic conversion of leaf chloroplasts into carotenoid-rich plastids reveals mechanistic basis of natural chromoplast development. *Proc. Natl Acad. Sci. USA*, **117**, 21796–21803.
- Lowder, L.G., Paul, J.W. and Qi, Y. (2017) Multiplexed transcriptional activation or repression in plants using CRISPR-dCas9-based systems. In *Plant Gene Regulatory Networks: Methods and Protocols* (Kaufmann, K. and Mueller-Roeber, B., eds), pp. 167–184. New York, NY: Springer.
- Lowder, L.G., Zhou, J., Zhang, Y., Malzahn, A., Zhong, Z., Hsieh, T.-F., Voytas, D.F. et al. (2018) Robust transcriptional activation in plants using multiplexed CRISPR-Act2.0 and mTALE-Act systems. *Mol. Plant*, **11**, 245–256.
- Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H. and Joung, J.K. (2013) CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods*, **10**, 977–979.
- Mahas, A., Aman, R. and Mahfouz, M. (2019) CRISPR-Cas13d mediates robust RNA virus interference in plants. *Genome Biol.* **20**, 263.
- Maioli, A., Gianoglio, S., Moglia, A., Acquadro, A., Valentino, D., Milani, A.M., Prohens, J. et al. (2020) Simultaneous CRISPR/Cas9 editing of three PPO genes reduces fruit flesh browning in *Solanum melongena* L. *Front. Plant Sci.* **11**, 1883.
- McKenzie, M.J., Mett, V., Reynolds, P.H.S. and Jameson, P.E. (1998) Controlled cytokinin production in transgenic tobacco using a copper-inducible promoter. *Plant Physiol.* **116**, 969–977.
- Misra, P., Pandey, A., Tiwari, M., Chandrashekar, K., Sidhu, O.P., Asif, M.H., Chakrabarty, D. et al. (2010) Modulation of transcriptome and metabolome of tobacco by Arabidopsis transcription factor, AtMYB12, leads to insect resistance. *Plant Physiol.* **152**, 2258–2268.
- Nabavi, S.M., Šamec, D., Tomczyk, M., Milella, L., Russo, D., Habtemariam, S., Sutar, I. et al. (2020) Flavonoid biosynthetic pathways in plants: versatile targets for metabolic engineering. *Biotechnol. Adv.* **38**, 107316.
- Naing, A.H., Park, K.I., Ai, T.N., Chung, M.Y., Han, J.S., Kang, Y.-W., Lim, K.B. et al. (2017) Overexpression of snapdragon Delila (Del) gene in tobacco enhances anthocyanin accumulation and abiotic stress tolerance. *BMC Plant Biol.* **17**, 65.
- Neelam, Khatkar, A. and Sharma, K.K. (2020) Phenylpropanoids and its derivatives: biological activities and its role in food, pharmaceutical and cosmetic industries. *Crit. Rev. Food Sci. Nutr.* **60**, 2655–2675.
- Norkunas, K., Harding, R., Dale, J. and Dugdale, B. (2018) Improving agroinfiltration-based transient gene expression in *Nicotiana benthamiana*. *Plant Methods*, **14**, 71.
- Ochoa-Fernandez, R., Abel, N.B., Wieland, F.-G., Schlegel, J., Koch, L.-A., Miller, J.B., Engesser, R. et al. (2020) Optogenetic control of gene expression in plants in the presence of ambient white light. *Nat. Methods*, **17**, 717–725.
- Ochoa-Fernandez, R., Samodelov, S.L., Brandl, S.M., Wehinger, E., Müller, K., Weber, W. and Zurbriggen, M.D. (2016) Optogenetics in plants: Red/Far-red light control of gene expression. *Methods Mol. Biol.* **1408**, 125–139.
- Pan, C., Wu, X., Markel, K., Malzahn, A.A., Kundagrami, N., Sretenovic, S., Zhang, Y. et al. (2021) CRISPR-Act3.0 for highly efficient multiplexed gene activation in plants. *Nat. Plants*, **7**, 942–953.
- Papikian, A., Liu, W., Gallego-Bartolomé, J. and Jacobsen, S.E. (2019) Site-specific manipulation of Arabidopsis loci using CRISPR-Cas9 SunTag systems. *Nat. Commun.* **10**, 729.
- Park, C.H., Xu, H., Yeo, H.J., Park, Y.E., Hwang, G.-S., Park, N.I. and Park, S.U. (2021) Enhancement of the flavone contents of *Scutellaria baicalensis* hairy roots via metabolic engineering using maize Lc and Arabidopsis PAP1 transcription factors. *Metab. Eng.* **64**, 64–73.
- Park, J.-J., Dempewolf, E., Zhang, W. and Wang, Z.-Y. (2017) RNA-guided transcriptional activation via CRISPR/dCas9 mimics overexpression phenotypes in Arabidopsis. *PLoS One*, **12**, e0179410.
- Pasin, F., Bedoya, L.C., Bernabé-Orts, J.M., Gallo, A., Simón-Mateo, C., Orzaez, D. and García, J.A. (2017) Multiple T-DNA delivery to plants using novel mini binary vectors with compatible replication origins. *ACS Synth. Biol.* **6**, 1962–1968.
- Pastor, V., Sánchez-Bel, P., Gamir, J., Pozo, M.J. and Flors, V. (2018) Accurate and easy method for systemin quantification and examining metabolic changes under different endogenous levels. *Plant Methods*, **14**, 33.
- Platek, A., Ali, Z., Baazim, H., Li, L., Abulfaraj, A., Al-Shareef, S., Aouida, M. et al. (2015) RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors. *Plant Biotechnol. J.* **13**, 578–589.
- Randall, R.S. (2021) The plant AlCR-pAlcA ethanol-inducible system displays gross growth artefacts independently of downstream pAlcA-regulated inducible constructs. *Sci. Rep.* **11**, 2142.
- Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J. et al. (2013) GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.* **162**, 1618–1631.
- Schymanski, E.L., Gerlich, M., Ruttkies, C. and Neumann, S. (2014) Solving CASMI 2013 with MetFrag, MetFusion and MOLGEN-MS/MS. *Mass Spectrom.* **3**(Spec Iss 2):S0036.
- Selma, S., Bernabé-Orts, J.M., Vazquez-Vilar, M., Diego-Martin, B., Ajenjo, M., Garcia-Carpintero, V., Granell, A. et al. (2019) Strong gene activation in plants with genome-wide specificity using a new orthogonal CRISPR/Cas9-based programmable transcriptional activator. *Plant Biotechnol. J.*, **17**, 1703–1705.
- Tang, X., Lowder, L.G., Zhang, T., Malzahn, A.A., Zheng, X., Voytas, D.F., Zhong, Z. et al. (2017) A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants*, **3**, 1–5.
- Tiwari, S.B., Belachew, A., Ma, S.F., Young, M., Ade, J., Shen, Y., Marion, C.M. et al. (2012) The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors: strong plant transcriptional activation domain. *Plant J.* **70**, 855–865.
- Torti, S., Schlesier, R., Thümmel, A., Bartels, D., Römer, P., Koch, B., Werner, S. et al. (2021) Transient reprogramming of crop plants for agronomic performance. *Nat. Plants*, **7**, 159–171.
- Vazquez-Vilar, M., Bernabé-Orts, J.M., Fernandez-del-Carmen, A., Ziarsolo, P., Blanca, J., Granell, A. and Orzaez, D. (2016) A modular toolbox for gRNA-Cas9 genome engineering in plants based on the GoldenBraid standard. *Plant Methods*, **12**, 10.
- Vu, A.T. and Lee, J.M. (2019) Genetic variations underlying anthocyanin accumulation in tomato fruits. *Euphytica*, **215**, 196.
- Wang, X. (2009) Structure, mechanism and engineering of plant natural product glycosyltransferases. *FEBS Lett.* **583**, 3303–3309.
- Xie, D.-Y., Sharma, S.B., Wright, E., Wang, Z.-Y. and Dixon, R.A. (2006) Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor. *Plant J.* **45**, 895–907.
- Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C. et al. (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, **160**, 339–350.
- Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., Kawai, P.G. et al. (2015) Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat. Commun.* **6**, 1–11.
- Zhao, L., Gao, L., Wang, H., Chen, X., Wang, Y., Yang, H., Wei, C. et al. (2013) The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis. *Funct. Integr. Genomics*, **13**, 75–98.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary Material Complete supplementary figures and tables captions.

Figure S1 gRNA optimization and selection.

Figure S2 Negative-ion mode fragmentation of naringenin.

Figure S3 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.

Table S1 List of candidate flavonoid genes identified in *N. benthamiana*

Table S2 Metabolite tentative identification in AP-N, AP- E, AP-K, and AP- Q samples

Table S3 List of GoldenBraid plasmids used in this work

Table S4 Primer pairs for RT-qPCR analyses