Title

A Suite of Constitutive Promoters for Tuning Gene Expression in Plants

Journal

ACS Synthetic Biology, 12(5)

ISSN

2161-5063

Authors

Zhou, Andy

Kirkpatrick, Liam D

Ornelas, Izaiah J

et al.

Publication Date

2023-05-19

DOI

10.1021/acssynbio.3c00075

A suite of constitutive promoters for tuning gene expression in plants

Authors: Andy Zhou[1,2,3], Liam D. Kirkpatrick[1,2,3], Izaiah J. Ornelas[2,3,4], Lorenzo J. Washington[1,2,3], Niklas F. C. Hummel[2,3], Christopher W. Gee[2,3], Sophia N. Tang[2,3,4], Collin R. Barnum[5], Henrik V. Scheller[1,2,3], Patrick M. Shih[1,2,3,6,*]

Affiliations:

[1] Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA 94720

[2] Feedstocks Division, Joint BioEnergy Institute, Emeryville, CA, USA 94608

[3] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA 94705

[4] Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA 94720

[5] Biochemistry, Molecular, Cellular and Developmental Biology Graduate Group, University of California, Davis, CA 95616

[6] Innovative Genomics Institute, University of California, Berkeley, CA 94720

[*] Corresponding Author. Email: pmshih@berkeley.edu

Abstract:

The need for convenient tools to express transgenes over a large dynamic range is pervasive throughout plant synthetic biology; however, current efforts are largely limited by the heavy reliance on a small set of strong promoters, precluding more nuanced and refined engineering endeavors. To address this technical gap, we characterize a suite of constitutive promoters that span a wide range of transcriptional levels and develop a GoldenGate-based plasmid toolkit named PCONS, optimized for versatile cloning and rapid testing of transgene expression at varying strengths. We demonstrate how easy access to a stepwise gradient of expression levels can be used for optimizing synthetic transcriptional systems and the production of small molecules *in planta*. We also systematically investigate the potential of using PCONS as an internal standard in plant biology experimental design, establishing best practices for signal normalization in experiments. Although

our library has primarily been developed for optimizing expression in *N. benthamiana*, we demonstrate the translatability of our promoters across distantly related species using a multiplexed reporter assay with barcoded transcripts. Our findings showcase the advantages of the PCONS library as an invaluable toolkit for plant synthetic biology.

Keywords: synthetic biology, plant, transgene expression, promoter library, STARR seq, transcriptome mining

Introduction:

Modulation of transgene expression is crucial for plant bioengineering applications such as constructing gene circuits, plant metabolic engineering, or increasing efficacy of gene editing efforts where the amount of protein being expressed by a synthetic promoter can drastically change outcomes [1–4]. Although a popular approach to tuning gene expression is the use of synthetic transcriptional systems, the lack of promoter choice limits engineering efforts to a single regime of universally high transgene expression strength and is an overall ineffective toolkit for modulating gene expression with a large dynamic range. The development of constitutive promoter libraries that drive a high dynamic range of expression strengths has been invaluable for synthetic biology efforts in most model microbial species. For example, the early development of such promoters in *Escherichia coli* through the systematic introduction of mutations on endogenous *cis*-elements has been widely utilized and improved the precision of experimental designs and metabolic engineering efforts in *E. coli* [5,6]. Thus, extending these approaches and the development of analogous parts would highly benefit the plant science community. Specifically, the model plant system *Nicotiana benthamiana* has emerged as an ideal platform for either rapidly prototyping designs or for biomanufacturing using advanced plant metabolic engineering [7–9]. Given the reliance on *N. benthamiana* by many plant synthetic biologists, the development of a suite of promoters with well-characterized gene expression strengths may provide the underlying tools to support a wide range of endeavors including pathway optimization, gene circuit designs, and genome engineering efforts.

There is significant interest in developing sophisticated genetic circuits in plants. Recent approaches for building gene circuits using either bacterial operators, *trans*-elements, or integrases fundamentally rely on plant promoter elements and precise expression strengths of these parts to ensure gene circuit function and to impart specificity to gene expression within the circuit [4,10]. One popular approach to tune gene expression is the use of orthogonal *trans*-elements, such as yeast Gal4 or a transcription-activation-like-effector (TALE), to target *cis*-elements to activate expression of a downstream transgene and has been demonstrated as an effective tool for transgenic protein expression in plant systems [3,11,12]. The complexities that afford tunability in such synthetic transcriptional systems may also lead to logistical challenges in construct design and implementation due to the requirement of introducing at least two components (e.g., the synthetic *trans*-element and synthetic promoter) to drive the expression of potentially only a single transgene. Thus for simple applications, it may instead be more practical to utilize a well-characterized promoter with a defined expression level. Furthermore, a fully quantified library of plant promoters can still be utilized as a modular component for complex gene circuit designs such as tuning expression of a synthetic *trans*-element or supplementing baseline expression of an inducible promoter [13,14]. Finally, the overuse of a single promoter can often lead to transgene silencing; thus, there is a need for alternative promoters that have similar expression levels that can be used in plant engineering efforts [15–17]. Such issues are made more prominent given the dearth of choice for plant constitutive promoters resulting in promoter designs primarily utilizing CaMV 35S enhancer, Nopaline synthase promoter variants, or ubiquitin and actin *cis*-regulatory elements[18–22]. Altogether, the development of a library of self-contained constitutive promoters that can express transgenes at various well-defined strengths would simplify construct design, enable faster pilot experiments in plant systems, and provide a versatile tool for plant synthetic biologists.

In this study, we developed a promoter library with well-defined gene expression strengths and consistent activity in multiple plant systems. Our library encompasses 15 promoters which span two orders of magnitude of expression strength with 5 promoters on par with 35S promoter activity in *N. benthamiana* transient expression assays. We extensively characterize the expression of transgenes driven by these

promoters in terms of biological variation and transgene identity. We evaluate the efficacy of using these promoters as internal standards for normalization of transgene expression in experimental contexts. Finally, we demonstrate that PCONS promoter activity is conserved in diverse plant species such as *Lactuca sativa* using self-transcribing active regulatory region sequencing (STARR-seq) and can drive transgene expression in *Medicago truncatula* via hairy-root transformation. Overall, these findings enable the adoption of the PCONS promoter library for use in any application where precise control of transgene expression strength is a necessity.

Results:

Mining and validation of a constitutive promoter library from publicly-available RNA-seq datasets

We envisioned the development of a plant promoter library with (1) high dynamic range, (2) ubiquitous expression in different tissues, and (3) activity in different plant species. With the abundance of RNA-seq datasets measuring Arabidopsis mRNA abundance spanning a wide array of tissues, organs, and developmental stages, we sought to integrate this information into a computational pipeline for extracting promoters displaying consistent expression activity across all tissues. In our promoter discovery pipeline, we emphasized filtering candidate promoters based on mean and variance gene expression quantiles across samples to identify and isolate promoters for a given plant transcriptomics dataset with consistent gene expression with respect to tissues and/or plant development (Figure 1A). We applied this pipeline to a widely-used Arabidopsis developmental transcriptomics atlas by Klepikova *et al* , which evaluated 79 organs and developmental stages of Arabidopsis, in order to demonstrate the efficacy of the pipeline to identify ubiquitously expressed promoters in different plant organs and developmental space and to build and validate a ready-to-use promoter library for many dicot systems[23]. We selected a subset of tissue samples from the Klepikova atlas which were largely uncorrelated in terms of expression of each individual gene (Figure S1) in order to maximize our candidate pool coverage of gene expression variance with respect to plant development and tissue type. By maximizing the gene expression variance space and subsequently filtering and constraining this space for low

variance promoters within the dataset, we enrich promoter designs which likely display consistent gene expression with respect to plant development and tissue type (Figure 2B). Additionally, these enriched promoters may contain an abundance of *cis*-regulatory elements tuned for constitutive gene expression which may be conserved and thus are more likely to display desirable activity in other plants. mRNA abundances obey Zipf's law, where a few genes account for the majority of gene expression [24]. Given this relationship between gene rank and mRNA abundance, we selected for the top 5 percent of genes with respect to mean expression which encompassed approximately 44% of all mapped reads in our compiled dataset thus enriching for a range of potential promoters with a wide but measurable dynamic range. In total, genes were subject to quantile filters selecting for high (>95 percentile) mean expression and low variance in expression (<30 percentile) resulting in 546 candidate genes (~2% of total genes) in Arabidopsis with sufficiently strong and spatially constitutive expression activity detectable in downstream characterization assays (Figure 1B).

To build a suite of constitutive promoters with a wide range of transgene expression strengths as a plant synthetic biology toolkit, we selected 15 genes spanning different mean abundance of transcripts from the enriched candidate pool (Table 1). We extracted a region of interest starting two kilobases upstream of the start codon to the beginning of the coding sequence of each gene to ensure capture of *cis*-regulatory elements responsible for their respective transcriptional activity [25–27]. Promoters were inserted into a pCAMBIA backbone upstream of PhytoBrick compatible GFP dropout cassette for facile construction of plasmid designs for metabolic engineering or plant synthetic biology (Figure 2A). This simple cloning strategy allows for the GFP dropout cassette to be replaced by the transgene of interest; thus, bacterial colonies with successful construction of the plasmid that have lost the GFP dropout cassette no longer fluoresce green and can be easily screened for on an agar plate. We sampled the transgene expression of these 15 promoters driving a reporter GFP and measured fluorescence intensity in *N. benthamiana* leaf disks after agroinfiltration. The full suite of promoters exhibited expression levels spanning nearly two orders of magnitude (Figure 2A). Notably, seven of our promoters achieve expression strengths that are stronger than the NOS promoter, and five are on par with the 35S promoter. From these GFP

measurements, we divided the 15 constitutive promoters into three groups denoted as low (PCL), medium (PCM), and high (PCH) and denoted the ensemble of promoters as the PCONS library. PCONS promoter strength measured through reporter gene expression was positively correlated (ρ=0.424) with Klepikova atlas RNA-seq readouts (Figure 2B, C).

Titratable activation of transgene expression using PCONS

Agroinfiltration in *N. benthamiana* is a gold standard methodology for plant metabolic engineering. One of the most popular plasmid systems used for *N. benthamiana*-based metabolic engineering efforts is the pEAQ plasmid system which uses 35S and cowpea mosaic virus UTRs to enable strong levels of transgene expression [22]. To further validate the dynamic range of the PCONS promoter library, we demonstrate distinct and incrementally increasing gene expression of the RUBY reporter system with 5 different PCONS promoters into the same *N. benthamiana* leaf [28]. We observed increasing red coloration of the infiltrated spots as PCONS promoter strength increased (Figure 3A). The gradation of betalain production observed in *N. benthamiana* leaf presents a direct demonstration of modulating metabolic flux, displays PCONS as distinct modules with consistent expression strengths for synthetic biology, and supports the usage of PCONS promoters as a potential tool for metabolic engineering.

Beyond metabolic engineering, the development of synthetic transcriptional systems and genetic circuits in plants has been of great interest. Orthogonal *trans*-elements, such as the yeast Gal4 transcription factor, are often utilized as a binary switch to toggle expression of a target gene and are the lynchpin component in recent efforts to build complex circuits in plants [4,11]. As such, the ability of *trans*-regulatory elements in activation of expression of a target gene is often viewed as strictly on/off depending on the presence or absence of the element in a given system. However, gene expression could also be conceptualized as a function where the rate of mRNA production depends on the concentration of TF input and thus further granularity of gene expression activation can be achieved by changing the expression level of a given *trans*-element through an upstream promoter [29]. *Trans*-elements have a concentration-dependent dynamic range in their ability to

activate gene expression, and within this range, tunable expression can be achieved exclusively through modulating *trans*-element concentration. To uncover the dynamic range of activation for a popular activating *trans*-element, Gal4:VP16, we used six promoters from our PCONS library to span a range of expression strengths to modulate its expression (Figure 3B). By measuring a downstream GFP reporter with Gal4 upstream activation sequences (UAS) as binding sites, we observe incremental activation of GFP with increasing expression of Gal4:VP16 (Figure 3C). Notably, the reporter signal is saturated when using medium to high promoter strengths (pCM2-pCH5). The dynamic range of GFP fluorescence is observed only when using weaker promoters (pCL). When these GFP readouts are framed with respect to previous PCONS measurements (Figure 2A) as a proxy for Gal4:VP16 concentration (Figure 3D), the data recapitulates a binding curve where the GFP readout saturates when nuclear Gal4:VP16 concentration is much higher than the Gal4 K_D. Thus, tuning the activation strength of Gal4:VP16 is achieved exclusively using weaker promoters within the PCONS library and further emphasizes the need for synthetic parts with diverse expression strengths necessary to access tunability and dynamics for a given transgene expression platform.

PCONS show consistent and correlated transgene expression of two fluorescent reporters

Shallow and deep machine learning models have demonstrated that gene coding sequences (CDS) and mRNA stabilities have strong predictive power on the overall expression of a given gene thus suggesting that the CDS context could potentially dictate promoter activity[30]. In order for the PCONS promoter library to be an effective toolkit for plant synthetic biology, strong PCONS promoters should remain strong relative to weaker PCONS promoters regardless of CDS content. Thus we aimed to establish whether the relative activity between different PCONS promoters can change due to transgene identity.

We tested this facet of PCONS through co-infiltration of two reporter fluorescent proteins (GFP and mScarlet) driven by the same promoter for each respective PCONS promoter. Visualizing all PCONS promoters individually driving two reporter

proteins in the same plot, we observed positive correlation between expression of GFP and mScarlet (Figure 4A, B). Each individual promoter displayed correlated expression strengths between mScarlet and GFP for both technical and biological replicates and formed clusters according to biological replicates (Figure 4A). These trends suggest that within a given single promoter, transgene expression strength is consistently maintained regardless of differences in transgene identity, plant state, and technical error for the entire PCONS library (Figure 4B). While the rank order of promoter strengths is not exactly the same when using PCONS to drive GFP versus mScarlet, promoter activity is positively correlated in expressing each transgene (ρ=0.862), although certain promoters may favor expression of one protein over another (Figure 4C). This correlated gene expression activity corroborates the robustness of PCONS promoters in driving expression of two fluorescent reporters.

Optimizing PCONS for signal normalization in plant gene expression studies

Normalization of signals is fundamental in the design of reproducible biological experiments by increasing precision of measurements and reducing certain facets of systematic and biological error affecting both the gene of interest and the normalizing gene. In the context of quantifying plant gene expression using agroinfiltration, many different promoters, reporters, and construct designs utilize dual luciferase or dual fluorescence reporters for signal normalization. Despite the popularity in the use of a second reporter to enable greater precision in readouts for promoter strength, empirically identifying ideal constructs for normalization has been less well-studied. Through evaluation of PCONS promoters as an internal standard we observed: (1) the reduction in the noisiness of signal quantified by the coefficient of variation, and (2) presence of a maximum capacity for transgene expression caused by using overtly strong PCONS promoters for signal normalization.

We evaluated the robustness of normalization by infiltrating individual PCONS promoters driving expression of GFP and co-infiltrating a choice between weak or strong PCONS promoters (PCL2 and PCH5 respectively) to drive expression of

mScarlet to normalize GFP fluorescence readouts. Replicate variation is spread out into different dynamic ranges of mScarlet signal depending on the choice of PCONS driving mScarlet expression (Figure 5A). Fluorescent readout was correlated between the mScarlet and GFP channels when using either low or high PCONS strength for normalization and with positive covariance overall between the two channels (Figure 5B) suggesting that each fluorescent channel captures a correlated biological variation, ultimately corrected through normalization, and thus boosting the resolution of differences between each individual PCONS promoter. To quantify the efficacy of normalization on reducing the experimental error, we determined the coefficient of variation (CV) or relative standard deviation of each PCONS promoter for both the raw GFP signal and normalized signal. After normalization, the coefficient of variation for each promoter generally decreases (CV ratio <1) demonstrating the efficacy of using PCONS promoters to normalize gene expression measurements (Figure 5C). In specific cases, the coefficient of variation increases after normalization, suggesting additional interaction between the promoter used for biological signal measurement and the promoter used for normalization. Therefore, the effect of normalizing on the coefficient of variation of the measured signal must be considered when designing experiments with internal standards. Overall, the reduced CV in normalized signal suggests that PCONS promoters can be an effective choice for an internal standard for gene expression studies.

In a wide range of microbial studies, resource allocation has been demonstrated to affect the physiology and overall metabolism of biological systems [31–33]. The introduction of strong exogenous promoter elements driving expression of transgenes imparts metabolic burden and limits cellular resources. While normalization with low or high PCONS promoters provides a reduction in measurement noise, we observed significant attenuation of the bulk GFP signal when using the strongest PCONS promoter (PCH5) for normalization versus a weaker promoter (PCL2) (Figure 5D). This finding suggests a maximum capacity for transgene expression in a plant system which can be a limiting factor depending on user experimental goals. Our results shed light on how transient expression studies in *N. benthamiana* that simultaneously express multiple genes at high expression levels may impact transcriptional levels and potentially the overall metabolic state

of transformed plant cells. These effects should be accounted for through the careful selection of promoters, highlighting the utility of the PCONS library.

PCONS library displays conserved activity in different dicot systems

To validate the general activity of the PCONS library in different plant systems, we quantified transcriptional activation of each promoter using a STARR-seq assay (Figure 6A) [34,35]. We generated a mixed plasmid library of PCONS promoters driving GFP or DsRed CDSs containing an 18nt barcode introduced via PCR. We subsequently agro-infiltrated the library into *N. benthamiana* and *L. sativa,* and through next generation sequencing of the resultant barcoded RNA transcripts, obtained simultaneous readout of promoter activity in driving expression of each transgene. Fold-enrichment of both GFP and DsRed transcripts driven by PCONS promoters share similar dynamic range in both *N. benthamiana* and *L. sativa* (Fig 6B). Furthermore, we observed a positive correlation in transcriptional activation of GFP and DsRed between both plant systems, suggesting that PCONS promoter activity is conserved despite millions of years of evolutionary divergence between each clade (Figure 6C). The GFP transcript abundance for each PCONS promoter is generally consistent with protein level measurements based on fluorescence measurements (Figure 6D). The 35S promoter displays the strongest transcriptional activation activity despite lower fluorescence readouts when compared with high activity PCONS promoters (Figure 2A). The high transcription activation strength of 35S could potentially incur a detrimental consequence on production of fluorescent protein aligned with our observation of an attenuation of fluorescence readout when using strong promoters for both readout and normalization (Figure 6D) whereas usage of a single strong PCONS promoter may not yet incur this penalty.

Additionally, we wanted to confirm that the conserved PCONS activities observed in transient assays are maintained in stable integrations. Thus, we generated stable integrations in *A. thaliana* via Agrobacterium floral dipping for 5 different promoters in the PCONS library driving a GFP-GUS fusion protein. As expected, beta-glucuronidase (GUS) activity was visualized throughout transgenic Arabidopsis seedling root, stem, and leaves (Figure 7A). Conserved promoter activity across different plant systems is a common feature of popular promoters used for plant

synthetic biology. To explore the efficacy of the PCONS library in driving expression in another dicot, we performed hairy root transformation of *Medicago truncatula* to express DsRed driven by PCONS promoters (Figure 7B). Transformed roots expressed DsRed with levels comparable to 35S promoter though, visually, 35S expressed DsRed strongest in Medicago. These results suggest that PCONS promoter activities, mined from Arabidopsis transcriptomics, can translate to other dicot systems such as *M. truncatula*, *N. benthamiana, and L. sativa*. Thus the selected PCONS promoters demonstrate a potential to increase the promoter options for stable plant transformation and can circumvent potential gene silencing incurred by using the same promoter to express multiple transgenes.

Discussion:

New high-throughput approaches in plants have resulted in an unprecedented amount of potential parts for plant synthetic biology. Validation of gene expression at the protein level, benchmarking newly discovered parts in relation to standardized parts, and expanding the dynamic range of biological activity are crucial for enabling successful experimental designs and engineering efforts *in planta*. We present a library of plant promoters quantified at both RNA and protein level in the expression of a variety of transgenes compared directly to popular gold-standard plant constitutive promoters. We demonstrate the utility of the PCONS library in incrementally increasing expression of biosynthetic enzymes for betalain synthesis and tunable gene activation using Gal4:VP16. We dimensionalize technical and biological variation and quantify the efficacy of normalization using different PCONS promoters, thus demonstrating their potential as internal standards for normalizing biological variation. PCONS promoters displayed correlated transcriptional activity in both *L. sativa* and *N. benthamiana* and stable expression of GUS and DsRed in *A.thaliana* and *M. truncatula* respectively, demonstrating their widespread activity in multiple plant systems. Overall, PCONS promoter activity was enriched *in silico* for constitutive gene expression through tissues in Arabidopsis development using appropriate transcriptomics data which and these findings bolster the benefits of transcriptomic screening for engineering features in plant promoters[13]. However, a full empirical demonstration of constitutive activity of the PCONS library over plant development has yet to be demonstrated.

The dynamic range of transgene expression provides potential limits to the tunability for a given parts library. An exhaustive transcriptional assessment of all Arabidopsis core promoter elements using a massively-parallel reporter assay (MPRA) spanned approximately a 2^9-fold change at the transcript level[36]. In this study, the PCONS library displays a 2^6-fold enrichment of transcripts and protein between weakest and strongest PCONS promoters. Given that the PCONS promoters encompass the top 5% of expressed genes on average in *A. thaliana,* unenriched endogenous promoter sequences are likely to be weaker in expression strength. Thus, the dynamic range of PCONS could expand with the inclusion of promoters extracted from more weakly expressed genes; however, gene expression of these promoters may be harder to quantify if sufficiently weak in strength, and their utility in driving expression of transgenes may be limited. Ultimately, our data pipeline utilizing the Klepikova Arabidopsis development RNA-seq dataset was able to mine promoters which demonstrate consistent and generalizable performance in two unique dicot systems and the highly-detectable dynamic range in PCONS promoters could provide utility as candidate reference genes for quantitative PCR in Arabidopsis We envision our mining approach could apply analogously to other transcriptomics-rich plant systems such as rice and maize to further expand on the available compatible synthetic promoter scaffolds for engineering monocots.

Machine learning models for gene expression in yeast suggest that different *cis*-regulatory components can each have unique contributory effects on the overall gene output [30]. Indeed, successful strategies in tuning gene expression for plant synthetic biology take advantage of these regulatory facets by modularizing these cis-elements to achieve enhanced dynamic range and fine-tuned levels of transgene expression[12,37]. High levels of gene expression, comparable with 35S, can be achieved by leveraging orthogonal trans-element activators and patterning plant cis-regulatory elements into synthetic promoters. For the design strategy of the PCONS library, we took a sequence agnostic approach to extract 2kb upstream of the start codon of each gene regulated by a respective PCONS promoter. Therefore, the granular cis-regulatory information which contributes to the overall promoter activity of PCONS, some of which lie in intergenic regions, still remains experimentally unexplored. Thus, we envision the PCONS library as a useful tool

that can be utilized in whole or piecemeal combinations with previously developed *cis*-regulatory parts and strategies to further enrich the existing suite of tools for plant synthetic biology.

Plants are capable of natural synthesis of a variety of complex compounds, and modulation of metabolic flux for these endeavors requires fine-tuned expression of multiple transgenes in order to maximize production of useful biological products. A common preliminary strategy in metabolic engineering efforts entail overexpression of multiple biosynthetic enzymes using strong promoters which may result in overaccumulation of intermediate compounds[38–40]. Our findings explore the downstream consequences of using strong promoters to drive multiple transgenes. Using two fluorescent reporters (GFP and mScarlet), we found our strongest PCONS promoter (PCH5) attenuated GFP fluorescence when tasked with simultaneous production of both transgenes. We speculate that this effect demonstrates a cellular limit in protein expression capacity when using strong promoters to drive multiple transgenes. To address these potential challenges, future metabolic engineering efforts may be able to utilize the PCONS library to tune metabolic flux, decrease overall transcriptomic load on plant cells, and circumvent potential engineering pitfalls. Overall, the PCONS library offers plant scientists a simple, streamlined plasmid expression system with a selection of well-defined promoter strengths for transgene expression.

Methods:

Data pipeline for identifying and analyzing promoters for the PCONS library

32 highly divergent samples from the Klepikova atlas were identified using sample correlation values from [23] (Table S1). Klepikova RNA-seq BAM files were obtained from CyVerse Data Commons (https://datacommons.cyverse.org/) and converted into count data using featureCounts [41]. Read data is normalized using the regular log (rlog) and variance stabilizing (VST) transformations (DESeq2[42]) and quantile filtered for >95%-tile mean gene expression and <30% variance of gene expression. Enriched genes from both transformations are combined and arranged

by their coefficient of variation (variance / mean) resulting in 546 unique genes. 15 genes of differing coefficient of variation were chosen for downstream evaluation and construction of the PCONS library. Bootstrapping of Pearson correlation coefficient was performed between Klepikova atlas trimmed mean of m-values (edgeR[43]) and measured GFP fluorescence of PCONS promoters from this study. Specifically, TMM values were calculated and aggregated for 32 samples in the Klepikova atlas for each gene regulated by a PCONS promoter. 18 random pairs of data; RNA-seq TMM and fluorescence, were sampled with replacement for each PCONS promoter for 10,000 iterations to a confidence interval for the Pearson correlation coefficient between the two datasets.

Plasmid construction

All constructs are based off of the pCAMBIA plasmid backbone. PCONS promoter sequences were PCR amplified from *A. thaliana* gDNA with and assembled into a pCAMBIA2301 backbone through Gibson assembly using NEBuilder HiFi Assembly Mix (NEB) and kit protocols (Table S2, S3). A GFP dropout cassette with common syntax BsaI sites was inserted downstream of each promoter resulting in pC(L/M/H)(1-5)-GFPdropout plasmid for convenient subsequent cloning steps [44]. Gal4:VP16, mScarlet, DsRed, and mEGFP were introduced into each PCONS-GFP dropout construct using Type IIS restriction cloning described in previous literature[45]. All constructs and plasmid maps are publicly available on the JBEI Inventory of Composable Elements (https://public-registry.jbei.org/) (Table S4).

Leaf punch assay

PCONS plasmids were transformed into GV3101 *Agrobacterium tumefaciens* via electroporation. Agroinfiltration protocol was adapted from Sparkes et al [46]. Transformed agrobacterium were grown in LB liquid media with 50 μg/ml kanamycin, 50 μg/ml rifampicin, and 30 μg/ml gentamicin to between optical density (OD) 0.6 and 1 before diluting to OD 0.5 in agroinfiltration buffer (10 mm MgCl2, 10 mm MES, pH 5.6). *N. benthamiana* plants were grown and maintained in a temperature-controlled growth room at 25 °C and 60% humidity in 16/8-h light/dark cycles with a daytime PPFD of ~120 μmol/m^2s. *N. benthamiana* were germinated and grown in Sungro Sunshine Mix #4 supplemented with ICL Osmocote

14-14-14 fertilizer at 5mL/L and agroinfiltrated at 29 days of age. Constructs of interest were infiltrated into the fourth leaf (counting down from the top of the tobacco plant) and harvested 3 days post-infiltration. Eight leaf disks per plant were extracted and arranged atop 250μL of water in black, clear-bottom, 96-well microtiter plates (Corning). Measurements of GFP and RFP fluorescence were recorded for each leaf disk using a BioTek Synergy H1 microplate reader (Agilent). Technical replicates were uniformly rank filtered to the top 6 of 8 disk readouts for all conditions to eliminate empty leaf disk readouts. The code used for analysis of fluorescence data is available at (https://github.com/shih-lab/PCONS_Analysis).

STARR-seq library design

PCONS STARR-seq library was generated using the pPSup (https://www.addgene.org/149416/) plasmid as template replacing the coding sequence, and terminator regions with GFP or DSRed, and tUBQ3 terminator respectively. The resultant plasmid was barcoded through PCR followed by introduction of each PCONS using Golden Gate assembly and transformed into *E.coli* (XL1-Blue). Approximately 50 *E.coli* colonies of each pCONS construct harboring GFP or DsRed were mixed into a single pooled library. Plasmids were isolated for barcode association using PacBio Sequel II sequencing (Azenta). The complete PCONS library was subsequently transformed into *A. tumefaciens* (GV3101) harboring the pSoup helper plasmid through electroporation [47].

STARR-seq assay

30mL of Agrobacterium harboring the PCONS promoter library was grown to an OD of 1.0 and subsequently 5mL was set aside as an input sample after plasmid isolation with a Plasmid Plus Midi Kit (Qiagen). The PCONS promoter library was introduced into six independent tobacco and lettuce plants through leaf tissue using agroinfiltration. Tissues were subject to RNA extraction 3 days post-infiltration using an RNEasy Plant Mini Kit (Qiagen). Two technical replicates for cDNA synthesis were performed for each RNA sample using complementary primers specific for either GFP or DsRed using SuperScript™ IV Reverse Transcriptase (Thermo-Fisher). Barcoded forward primers (Table S2) were used for cDNA amplification for downstream demultiplexing of biological replicates after optimal amplification cycles were identified using qPCR (SYBR® Green Master Mix, BioRad). Sequencing

adapter ligation, TapeStation, and Bioanalyzer quality control of cDNA amplicons and 2x 150 paired-end sequencing was performed by Azenta life sciences using an Illumina HiSeq 4000.

Analysis of STARR-seq

In order to identify PCONS promoters from barcoded reads, a dictionary between PCONS promoters and barcodes was constructed from PacBio Sequel II sequencing of the input PCONS promoter library. Subreads were processed using the PacBio ccs package into an unaligned BAM file and subsequently aligned with PCONS promoter sequences. The aligned read barcodes and PCONS identities were then processed into a dictionary object in Python for downstream analysis. Paired-end sequencing data from plant cDNA barcoded amplicons were in mixed orientation due to ligation of sequencing adapters. Paired-end reads were combined using PANDAseq [48]. Cutadapt was used for adapter trimming and to reorient and demultiplex reads into their respective biological replicate and transgene identities [49]. Barcodes with less than five reads were discarded for downstream analysis. Each barcode frequency (barcode count divided by all counts) was mapped to their respective promoter identities, and enrichment of each barcode was calculated by dividing by the barcode frequency in the input sample. The average enrichment of all promoters was then calculated and visualized for all samples. Scripts used for processing and analysis of sequencing data are available at (https://github.com/shih-lab/PCONS_Analysis).

GUS expression in Arabidopsis seedlings

All *A. thaliana* were germinated and grown in Sungro Sunshine Mix #1 soil in a Percival growth chamber at 22°C and 60% humidity with short day (8/16-h light/dark) or long day (16/8-h light/dark) conditions at a daytime PPFD of ~200 µmol/m^2s. *A. thaliana* stably expressing a GFP-GUS fusion protein driven by various PCONS promoters were generated by floral dipping of healthy Col-0 plants in PCONS binary vector-harboring GV3101 *A. tumefaciens* and subsequent seed selection with kanamycin on ½ strength Murashige and Skoog (MS) agar media. T2 seeds were germinated and GUS expression was visualized after 10 days. GUS expression was visualized using an adjusted method based on prior established protocols [50]. 10-day-old Arabidopsis seedlings were submerged in staining solution consisting of 0.5

mM $K_3Fe(CN)_6$, 0.5 mM $K_4Fe(CN)_6$, 100 mM sodium phosphate buffer pH 7.0, 10 mM ethylenediaminetetraacetic acid, 0.5 mg/ml X-Gluc (5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid cyclohexylammonium salt) for 24 h at room temperature. Stained seedlings were sequentially washed with 70% and 100% ethanol for 8 h each. As a control, wildtype *A. thaliana* seedlings (Columbia ecotype) were studied using the same procedure.

Hairy root transformation

Medicago truncatula seeds were scarified in sulfuric acid and surface sterilized with 5% v/v commercial bleach with deionized water before being plated on ½ strength Murashige and Skoog (MS) agar media. After two days of stratification at 4°C in the dark followed by two days of incubation at 22°C, seedling radicles were removed with a scalpel and the wound inoculated with *Agrobacterium fabrum* ARqua1 harboring binary vectors containing PCONS promoters driving DsRed fluorescent protein expression [51]. Inoculated seedlings were co-cultured on ½ strength MS agar media for 5 days then selected on ½ strength MS agar media with kanamycin (25 micrograms / mL) and 0.1% v/v Plant Preservative Mixture. Images were taken 3 weeks after inoculation.

Microscopy

Imaging of stable integration Arabidopsis GUS seedlings and *Medicago* DsRed seedlings was performed on an Olympus SZX16 stereoscope fitted with an Olympus SZX2-ILLT base for transmitted/oblique illumination and a fluorescence filter wheel with X-Cite 120 excitation light source for fluorescence. Filters (Semrock BrightLine) for 562 nm (40 nm bandwidth) and 624 nm (40 nm bandwidth) were used for excitation of DsRed and emission filtering, respectively. Images were collected with a Lumenera INFINITY 2-5 color camera (gain = 20, exposure = 400 ms) with 0.63x adapter.

Abbreviations

STARR-seq: self-transcribing active regulatory region sequencing
PCC: Pearson correlation coefficient
UAS: upstream activating sequence
CDS: coding sequence

CV: coefficient of variation
GUS: β-glucuronidase gene, *uidA*
VST: variance stabilizing transformation
OD: optical density
MS: Murashige and Skoog medium

Accession Numbers

All barcode sequencing reads were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under the BioProject accession PRJNA922233 (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA922233/).

Author Contributions

A.Z and P.M.S. conceived and interpreted experiments and wrote the article; A.Z., L.D.K., I.J.O., L.J.W., N.F.C.H., C.W.G., S.N.T., C.R.B., performed experiments; and AZ analyzed the data and prepared the figures; P.M.S. and H.V.S. provided supervision of research.

Acknowledgements

We thank Dr. Simon Alamos for input and feedback on the manuscript. This work was supported by the DOE Joint BioEnergy Institute, which is supported by the Office of Science, Office of Biological and Environmental Research, the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. A.Z. was supported by NSF Postdoctoral Fellowship in Biology, Award Number 2009093. Figures 2 and 5 were created in part with Biorender.com

Conflict of Interest Statement

The authors declare no competing financial interest.

Supporting information

Supporting information contains details and accession IDs of Klepikova Arabidopsis developmental atlas samples used for mining PCONS promoters. PCONS sequences, primers, and plasmid information are included.

References:

(1) Gil-Humanes, J.; Wang, Y.; Liang, Z.; Shan, Q.; Ozuna, C. V.; Sánchez-León, S.; Baltes, N. J.; Starker, C.; Barro, F.; Gao, C.; Voytas, D. F. High-Efficiency Gene Targeting in Hexaploid Wheat Using DNA Replicons and CRISPR/Cas9. *Plant J.* **2017**, *89* (6), 1251–1262. https://doi.org/10.1111/tpj.13446.
(2) Dahan-Meir, T.; Filler-Hayut, S.; Melamed-Bessudo, C.; Bocobza, S.; Czosnek, H.; Aharoni, A.; Levy, A. A. Efficient in Planta Gene Targeting in Tomato Using Geminiviral Replicons and the CRISPR/Cas9 System. *Plant J.* **2018**, *95* (1), 5–16. https://doi.org/10.1111/tpj.13932.
(3) Brückner, K.; Schäfer, P.; Weber, E.; Grützner, R.; Marillonnet, S.; Tissier, A. A Library of Synthetic Transcription Activator-like Effector-Activated Promoters for Coordinated Orthogonal Gene Expression in Plants. *Plant J.* **2015**, *82* (4), 707–716. https://doi.org/10.1111/tpj.12843.
(4) Brophy, J. A. N.; Magallon, K. J.; Duan, L.; Zhong, V.; Ramachandran, P.; Kniazev, K.; Dinneny, J. R. Synthetic Genetic Circuits as a Means of Reprogramming Plant Roots. *Science* **2022**, *377* (6607), 747–751. https://doi.org/10.1126/science.abo4326.
(5) Kosuri, S.; Goodman, D. B.; Cambray, G.; Mutalik, V. K.; Gao, Y.; Arkin, A. P.; Endy, D.; Church, G. M. Composability of Regulatory Sequences Controlling Transcription and Translation in *Escherichia Coli*. *Proc. Natl. Acad. Sci.* **2013**, *110* (34), 14024–14029. https://doi.org/10.1073/pnas.1301301110.
(6) Mutalik, V. K.; Guimaraes, J. C.; Cambray, G.; Lam, C.; Christoffersen, M. J.; Mai, Q.-A.; Tran, A. B.; Paull, M.; Keasling, J. D.; Arkin, A. P.; Endy, D. Precise and Reliable Gene Expression via Standard Transcription and Translation Initiation Elements. *Nat. Methods* **2013**, *10* (4), 354–360. https://doi.org/10.1038/nmeth.2404.
(7) Schultz, B. J.; Kim, S. Y.; Lau, W.; Sattely, E. S. Total Biosynthesis for Milligram-Scale Production of Etoposide Intermediates in a Plant Chassis. *J. Am. Chem. Soc.* **2019**, *141* (49), 19231–19235. https://doi.org/10.1021/jacs.9b10717.
(8) Marsian, J.; Lomonossoff, G. P. Molecular Pharming—VLPs Made in Plants. *Curr. Opin. Biotechnol.* **2016**, *37*, 201–206. https://doi.org/10.1016/j.copbio.2015.12.007.
(9) Reed, J.; Stephenson, M. J.; Miettinen, K.; Brouwer, B.; Leveau, A.; Brett, P.; Goss, R. J. M.; Goossens, A.; O'Connell, M. A.; Osbourn, A. A Translational Synthetic Biology Platform for Rapid Access to Gram-Scale Quantities of Novel Drug-like Molecules. *Metab. Eng.* **2017**, *42*, 185–193. https://doi.org/10.1016/j.ymben.2017.06.012.
(10) Lloyd, J. P. B.; Ly, F.; Gong, P.; Pflueger, J.; Swain, T.; Pflueger, C.; Fourie, E.; Khan, M. A.; Kidd, B. N.; Lister, R. Synthetic Memory Circuits for Stable Cell Reprogramming in Plants. *Nat. Biotechnol.* **2022**, 1–11. https://doi.org/10.1038/s41587-022-01383-2.
(11) Belcher, M. S.; Vuu, K. M.; Zhou, A.; Mansoori, N.; Agosto-Ramos, A.; Loqué, D.; Shih, P. M. Design and Characterization of Orthogonal Regulatory Systems to Control Gene Expression in Plants. *Nat. Chem. Biol.* **2019**.
(12) Schaumberg, K. A.; Antunes, M. S.; Kassaw, T. K.; Xu, W.; Zalewski, C. S.; Medford, J. I.; Prasad, A. Quantitative Characterization of Genetic Parts and Circuits for Plant Synthetic Biology. *Nat. Methods* **2016**, *13* (1), 94–100. https://doi.org/10.1038/nmeth.3659.
(13) Jameel, A.; Noman, M.; Liu, W.; Ahmad, N.; Wang, F.; Li, X.; Li, H. Tinkering Cis Motifs Jigsaw Puzzle Led to Root-Specific Drought-Inducible Novel Synthetic Promoters. *Int. J. Mol. Sci.* **2020**, *21* (4), 1357.

https://doi.org/10.3390/ijms21041357.
(14) Zuo, J.; Chua, N.-H. Chemical-Inducible Systems for Regulated Expression of Plant Genes. *Curr. Opin. Biotechnol.* **2000**, *11* (2), 146–151. https://doi.org/10.1016/S0958-1669(00)00073-2.
(15) Martínez de Alba, A. E.; Elvira-Matelot, E.; Vaucheret, H. Gene Silencing in Plants: A Diversity of Pathways. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **2013**, *1829* (12), 1300–1308. https://doi.org/10.1016/j.bbagrm.2013.10.005.
(16) Meyer, P.; Saedler, H. Homology-Dependent Gene Silencing in Plants. *Annu. Rev. Plant Biol.* **1996**, *47* (1), 23–48.
(17) Rajeevkumar, S.; Anunanthini, P.; Sathishkumar, R. Epigenetic Silencing in Transgenic Plants. *Front. Plant Sci.* **2015**, *6*. https://doi.org/10.3389/fpls.2015.00693.
(18) Harpster, M. H.; Townsend, J. A.; Jones, J. D.; Bedbrook, J.; Dunsmuir, P. Relative Strengths of the 35S Cauliflower Mosaic Virus, 1', 2', and Nopaline Synthase Promoters in Transformed Tobacco Sugarbeet and Oilseed Rape Callus Tissue. *Mol. Gen. Genet. MGG* **1988**, *212* (1), 182–190. https://doi.org/10.1007/BF00322463.
(19) Huang, S.; An, Y.-Q.; McDowell, J. M.; McKinney, E. C.; Meagher, R. B. The Arabidopsis ACT11 Actin Gene Is Strongly Expressed in Tissues of the Emerging Inflorescence, Pollen, and Developing Ovules. *Plant Mol. Biol.* **1997**, *33* (1), 125–139. https://doi.org/10.1023/A:1005741514764.
(20) Christensen, A. H.; Quail, P. H. Ubiquitin Promoter-Based Vectors for High-Level Expression of Selectable and/or Screenable Marker Genes in Monocotyledonous Plants. *Transgenic Res.* **1996**, *5* (3), 213–218. https://doi.org/10.1007/BF01969712.
(21) Sanders, P. R.; Winter, J. A.; Barnason, A. R.; Rogers, S. G.; Fraley, R. T. Comparison of Cauliflower Mosaic Virus 35S and Nopaline Synthase Promoters in Transgenic Plants. *Nucleic Acids Res.* **1987**, *15* (4), 1543–1558. https://doi.org/10.1093/nar/15.4.1543.
(22) Sainsbury, F.; Thuenemann, E. C.; Lomonossoff, G. P. PEAQ: Versatile Expression Vectors for Easy and Quick Transient Expression of Heterologous Proteins in Plants. *Plant Biotechnol. J.* **2009**, *7* (7), 682–693. https://doi.org/10.1111/j.1467-7652.2009.00434.x.
(23) Klepikova, A. V.; Kasianov, A. S.; Gerasimov, E. S.; Logacheva, M. D.; Penin, A. A. A High Resolution Map of the Arabidopsis Thaliana Developmental Transcriptome Based on RNA-Seq Profiling. *Plant J. Cell Mol. Biol.* **2016**, *88* (6), 1058–1070. https://doi.org/10.1111/tpj.13312.
(24) Furusawa, C.; Kaneko, K. Zipf's Law in Gene Expression. *Phys. Rev. Lett.* **2003**, *90* (8), 088102. https://doi.org/10.1103/PhysRevLett.90.088102.
(25) Conforte, A. J.; Guimarães-Dias, F.; Neves-Borges, A. C.; Bencke-Malato, M.; Felix-Whipps, D.; Alves-Ferreira, M. Isolation and Characterization of a Promoter Responsive to Salt, Osmotic and Dehydration Stresses in Soybean. *Genet. Mol. Biol.* **2017**, *40* (1 Suppl 1), 226–237. https://doi.org/10.1590/1678-4685-GMB-2016-0052.
(26) Sorkina, A.; Bardosh, G.; Liu, Y.-Z.; Fridman, I.; Schlizerman, L.; Zur, N.; Or, E.; Goldschmidt, E. E.; Blumwald, E.; Sadka, A. Isolation of a Citrus Promoter Specific for Reproductive Organs and Its Functional Analysis in Isolated Juice Sacs and Tomato. *Plant Cell Rep.* **2011**, *30* (9), 1627–1640. https://doi.org/10.1007/s00299-011-1073-3.

(27) Negi, N.; Khurana, P. A Salicylic Acid Inducible Mulberry WRKY Transcription Factor, MiWRKY53 Is Involved in Plant Defence Response. *Plant Cell Rep.* **2021**, *40* (11), 2151–2171. https://doi.org/10.1007/s00299-021-02710-8.
(28) He, Y.; Zhang, T.; Sun, H.; Zhan, H.; Zhao, Y. A Reporter for Noninvasively Monitoring Gene Expression and Plant Transformation. *Hortic. Res.* **2020**, *7* (1), 1–6. https://doi.org/10.1038/s41438-020-00390-1.
(29) Ricci-Tam, C.; Ben-Zion, I.; Wang, J.; Palme, J.; Li, A.; Savir, Y.; Springer, M. Decoupling Transcription Factor Expression and Activity Enables Dimmer Switch Gene Regulation. *Science* **2021**, *372* (6539), 292–295. https://doi.org/10.1126/science.aba7582.
(30) Zrimec, J.; Börlin, C. S.; Buric, F.; Muhammad, A. S.; Chen, R.; Siewers, V.; Verendel, V.; Nielsen, J.; Töpel, M.; Zelezniak, A. Deep Learning Suggests That Gene Expression Is Encoded in All Parts of a Co-Evolving Interacting Gene Regulatory Structure. *Nat. Commun.* **2020**, *11*, 6141. https://doi.org/10.1038/s41467-020-19921-4.
(31) Ceroni, F.; Algar, R.; Stan, G.-B.; Ellis, T. Quantifying Cellular Capacity Identifies Gene Expression Designs with Reduced Burden. *Nat. Methods* **2015**, *12* (5), 415–418. https://doi.org/10.1038/nmeth.3339.
(32) Basan, M.; Hui, S.; Okano, H.; Zhang, Z.; Shen, Y.; Williamson, J. R.; Hwa, T. Overflow Metabolism in Escherichia Coli Results from Efficient Proteome Allocation. *Nature* **2015**, *528* (7580), 99–104. https://doi.org/10.1038/nature15765.
(33) Li, G.-W.; Burkhardt, D.; Gross, C.; Weissman, J. S. Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell* **2014**, *157* (3), 624–635. https://doi.org/10.1016/j.cell.2014.02.033.
(34) Arnold, C. D.; Gerlach, D.; Stelzer, C.; Boryń, Ł. M.; Rath, M.; Stark, A. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-Seq. *Science* **2013**, *339* (6123), 1074–1077. https://doi.org/10.1126/science.1232542.
(35) Jores, T.; Tonnies, J.; Dorrity, M. W.; Cuperus, J. T.; Fields, S.; Queitsch, C. Identification of Plant Enhancers and Their Constituent Elements by STARR-Seq in Tobacco Leaves[OPEN]. *Plant Cell* **2020**, *32* (7), 2120–2131. https://doi.org/10.1105/tpc.20.00155.
(36) Jores, T.; Tonnies, J.; Wrightsman, T.; Buckler, E. S.; Cuperus, J. T.; Fields, S.; Queitsch, C. Synthetic Promoter Designs Enabled by a Comprehensive Analysis of Plant Core Promoters. *Nat. Plants* **2021**, *7* (6), 842–855. https://doi.org/10.1038/s41477-021-00932-y.
(37) Cai, Y.-M.; Kallam, K.; Tidd, H.; Gendarini, G.; Salzman, A.; Patron, N. J. Rational Design of Minimal Synthetic Promoters for Plants. *Nucleic Acids Res.* **2020**, *48* (21), 11845–11856. https://doi.org/10.1093/nar/gkaa682.
(38) Lee, M. E.; Aswani, A.; Han, A. S.; Tomlin, C. J.; Dueber, J. E. Expression-Level Optimization of a Multi-Enzyme Pathway in the Absence of a High-Throughput Assay. *Nucleic Acids Res.* **2013**, *41* (22), 10668–10678. https://doi.org/10.1093/nar/gkt809.
(39) Ajikumar, P. K.; Xiao, W.-H.; Tyo, K. E. J.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T. H.; Pfeifer, B.; Stephanopoulos, G. Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia Coli. *Science* **2010**, *330* (6000), 70–74. https://doi.org/10.1126/science.1191652.
(40) Pfleger, B. F.; Pitera, D. J.; Smolke, C. D.; Keasling, J. D. Combinatorial Engineering of Intergenic Regions in Operons Tunes Expression of Multiple Genes. *Nat. Biotechnol.* **2006**, *24* (8), 1027–1032.

https://doi.org/10.1038/nbt1226.
(41) Liao, Y.; Smyth, G. K.; Shi, W. FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* **2014**, *30* (7), 923–930. https://doi.org/10.1093/bioinformatics/btt656.
(42) Love, M. I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15* (12), 550. https://doi.org/10.1186/s13059-014-0550-8.
(43) Robinson, M. D.; McCarthy, D. J.; Smyth, G. K. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26* (1), 139–140. https://doi.org/10.1093/bioinformatics/btp616.
(44) Patron, N. J.; Orzaez, D.; Marillonnet, S.; Warzecha, H.; Matthewman, C.; Youles, M.; Raitskin, O.; Leveau, A.; Farré, G.; Rogers, C.; Smith, A.; Hibberd, J.; Webb, A. A. R.; Locke, J.; Schornack, S.; Ajioka, J.; Baulcombe, D. C.; Zipfel, C.; Kamoun, S.; Jones, J. D. G.; Kuhn, H.; Robatzek, S.; Van Esse, H. P.; Sanders, D.; Oldroyd, G.; Martin, C.; Field, R.; O'Connor, S.; Fox, S.; Wulff, B.; Miller, B.; Breakspear, A.; Radhakrishnan, G.; Delaux, P.-M.; Loqué, D.; Granell, A.; Tissier, A.; Shih, P.; Brutnell, T. P.; Quick, W. P.; Rischer, H.; Fraser, P. D.; Aharoni, A.; Raines, C.; South, P. F.; Ané, J.-M.; Hamberger, B. R.; Langdale, J.; Stougaard, J.; Bouwmeester, H.; Udvardi, M.; Murray, J. A. H.; Ntoukakis, V.; Schäfer, P.; Denby, K.; Edwards, K. J.; Osbourn, A.; Haseloff, J. Standards for Plant Synthetic Biology: A Common Syntax for Exchange of DNA Parts. *New Phytol.* **2015**, *208* (1), 13–19. https://doi.org/10.1111/nph.13532.
(45) Patron, N. J. DNA Assembly for Plant Biology. *Curr. Protoc. Plant Biol.* **2016**, *1* (4), 604–616. https://doi.org/10.1002/cppb.20038.
(46) Sparkes, I. A.; Runions, J.; Kearns, A.; Hawes, C. Rapid, Transient Expression of Fluorescent Fusion Proteins in Tobacco Plants and Generation of Stably Transformed Plants. *Nat. Protoc.* **2006**, *1* (4), 2019–2025. https://doi.org/10.1038/nprot.2006.286.
(47) Hellens, R. P.; Edwards, E. A.; Leyland, N. R.; Bean, S.; Mullineaux, P. M. PGreen: A Versatile and Flexible Binary Ti Vector for Agrobacterium-Mediated Plant Transformation. *Plant Mol. Biol.* **2000**, *42* (6), 819–832. https://doi.org/10.1023/A:1006496308160.
(48) Masella, A. P.; Bartram, A. K.; Truszkowski, J. M.; Brown, D. G.; Neufeld, J. D. PANDAseq: Paired-End Assembler for Illumina Sequences. *BMC Bioinformatics* **2012**, *13* (1), 31. https://doi.org/10.1186/1471-2105-13-31.
(49) Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet.journal* **2011**, *17* (1), 10–12. https://doi.org/10.14806/ej.17.1.200.
(50) Jefferson, R. A.; Kavanagh, T. A.; Bevan, M. W. GUS Fusions: Beta-Glucuronidase as a Sensitive and Versatile Gene Fusion Marker in Higher Plants. *EMBO J.* **1987**, *6* (13), 3901–3907. https://doi.org/10.1002/j.1460-2075.1987.tb02730.x.
(51) Thompson, M. G.; Cruz-Morales, P.; Moore, W. M.; Pearson, A. N.; Keasling, J. D.; Scheller, H. V.; Shih, P. M. Draft Genome Sequence of Agrobacterium Fabrum ARqua1. *Microbiol. Resour. Announc.* **2020**, *9* (27), e00506-20. https://doi.org/10.1128/MRA.00506-20.

Figures:

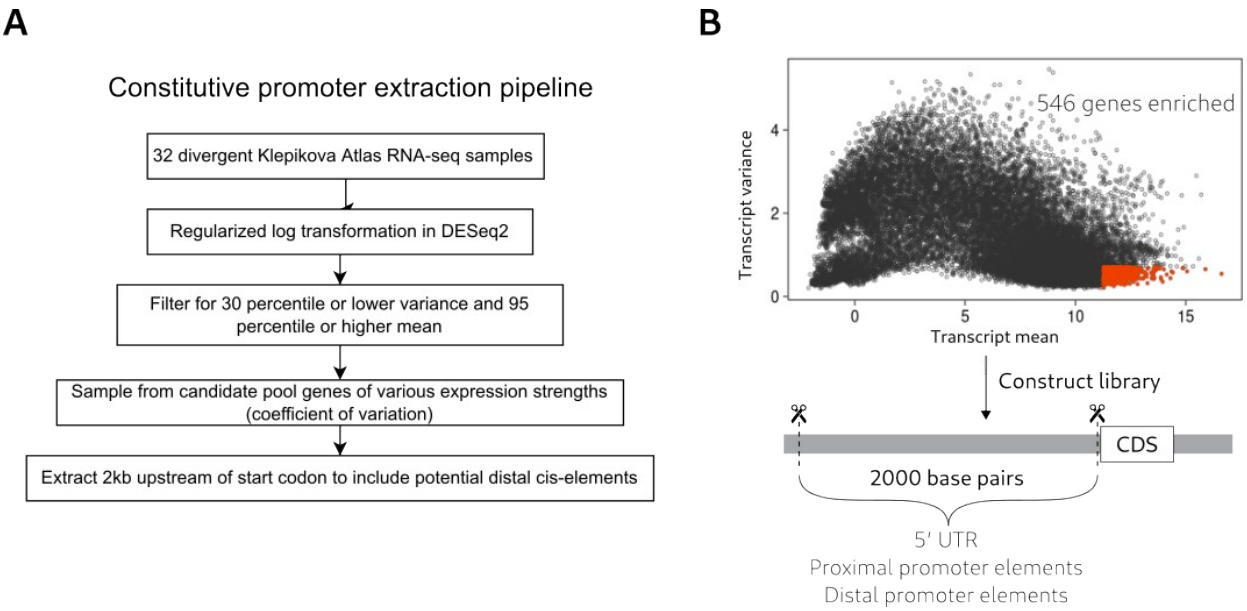

Figure 1. Data pipeline for identifying constitutive promoters from RNA-seq datasets and validation of promoter activity through transient expression in *N. benthamiana*. (A) Pipeline for extraction of constitutive promoters from the Klepikova RNA-seq dataset. (B) 32 regularized RNA-seq samples mapped by mRNA abundance mean and variance. Red highlighted regions of interest are quantile filtered genes by mean (>95%) and variance (<30%).

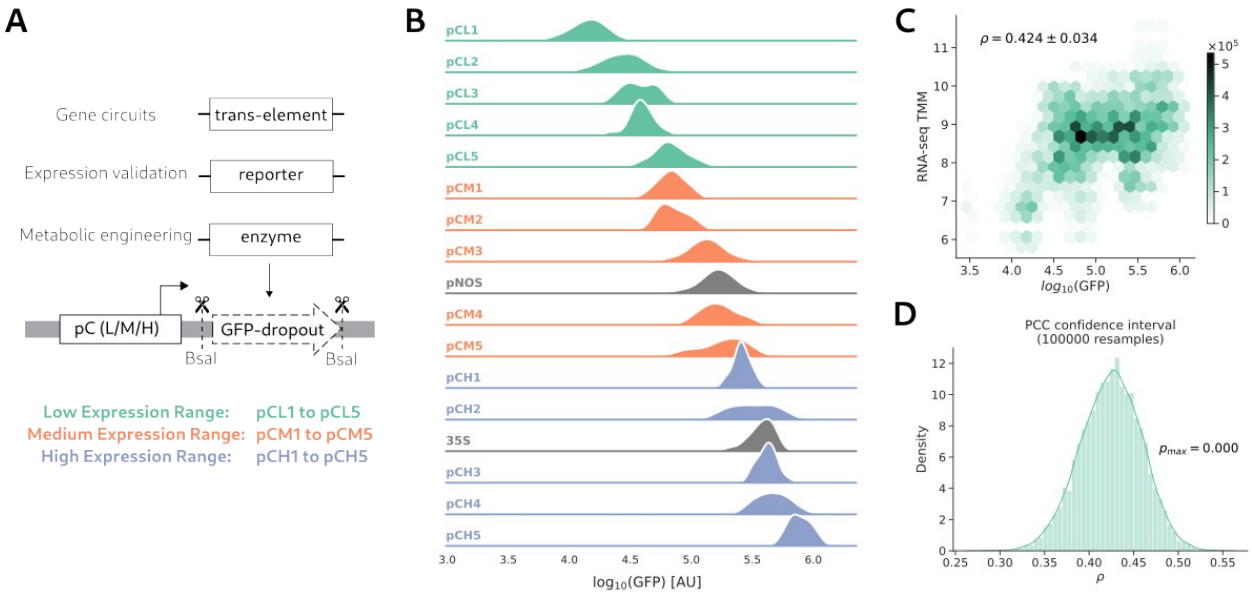

Figure 2. GFP expression output of PCONS promoters in agroinfiltrated *N. benthamiana* leaves and correlation with native *A. thaliana* mRNA abundances. (A) Design of GFP-dropout plasmid for convenient swapping of transgenes driven by a PCONS promoter. (B) Ridge plot displaying mEGFP expression of the PCONS library after agroinfiltration of constructs into *N. benthamiana* leaves. (C) Correlation between normalized mRNA abundance in TMM (trimmed mean of m-values) of PCONS genes and GFP expression driven by PCONS promoters. (D) Confidence interval of Pearson correlation coefficient (PCC) between GFP fluorescence and RNA-seq gene counts (as TMM) from the Klepikova RNA-seq dataset as extrapolated via bootstrapping.

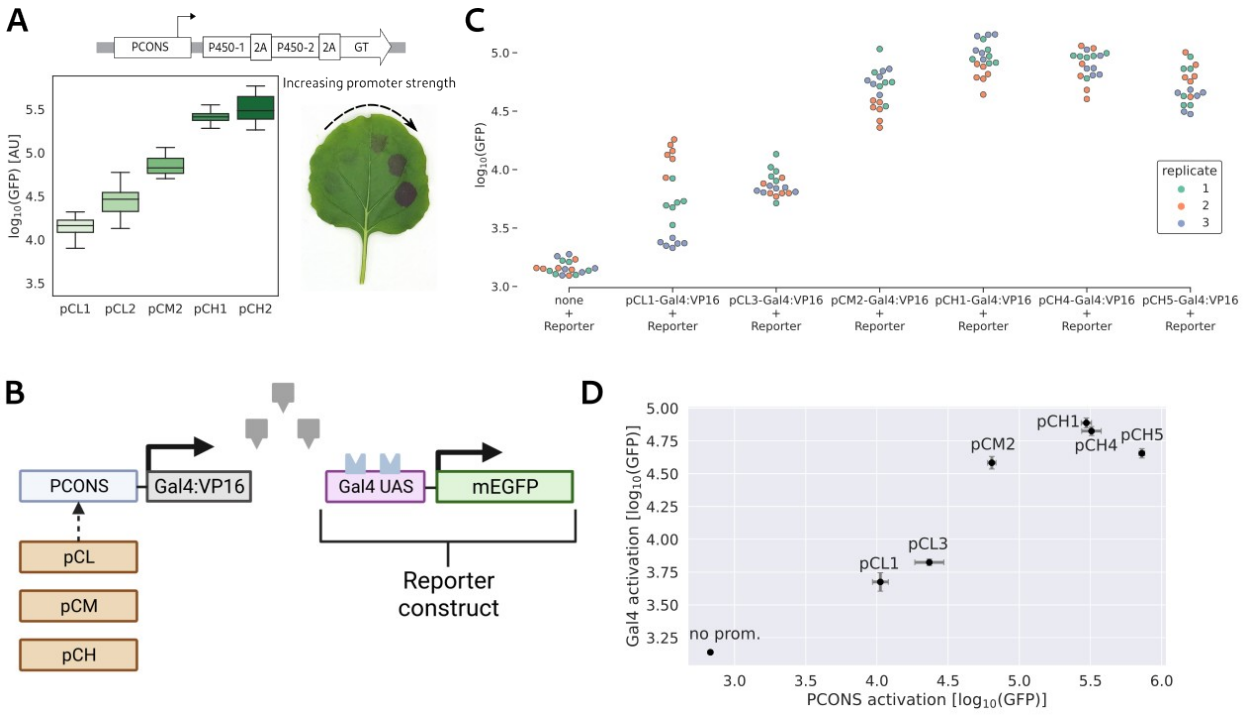

Figure 3. Stepwise expression of a metabolic pathway and *trans*-activator element using PCONS promoters. (A) Selection of 5 promoters used to express the betalain biosynthesis pathway with increasing strength in a single *N. benthamiana* leaf. Boxplots show GFP expression strengths of promoters used for betalain biosynthesis. Betalain is visualized but production was not quantified. (B) Visualization of mGFP expression driven by a Gal4 UAS using PCONS for regulation of Gal4:VP16 and (C) measurements using different PCONS upstream of Gal4:VP16. (D) Comparison of average fluorescence readout driven directly through PCONS promoters versus indirectly driven through expression of Gal4:VP16. Error bars represent standard error of mean.

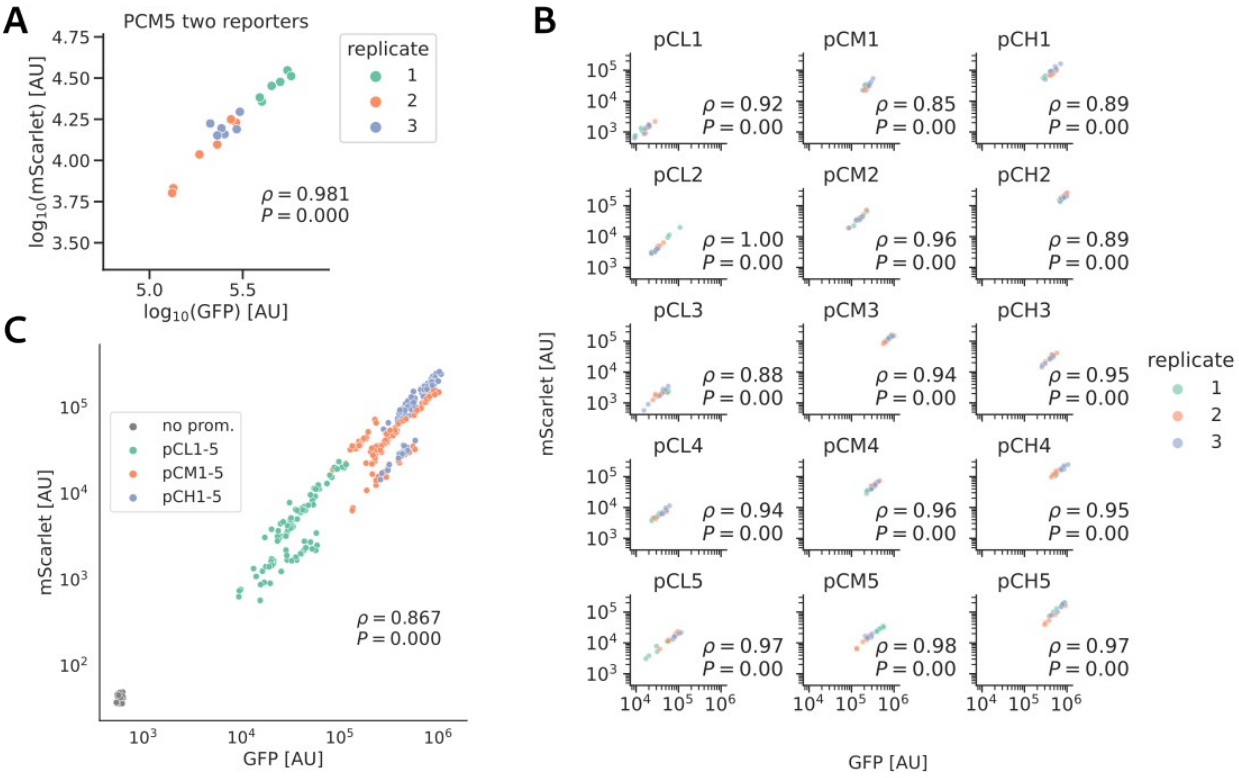

Figure 4. PCONS driven transgene expression bias and variation tracked through two fluorescent reporters. (A) Expression of mEGFP and mScarlet was simultaneously measured after co-infiltration of two constructs containing PCM5 promoter driving expression of each fluorophore. (B) Co-infiltration of two fluorescent reporters driven by each PCONS promoter. (C) Correlation of PCONS promoter activity when simultaneously driving two different fluorescent transgenes.

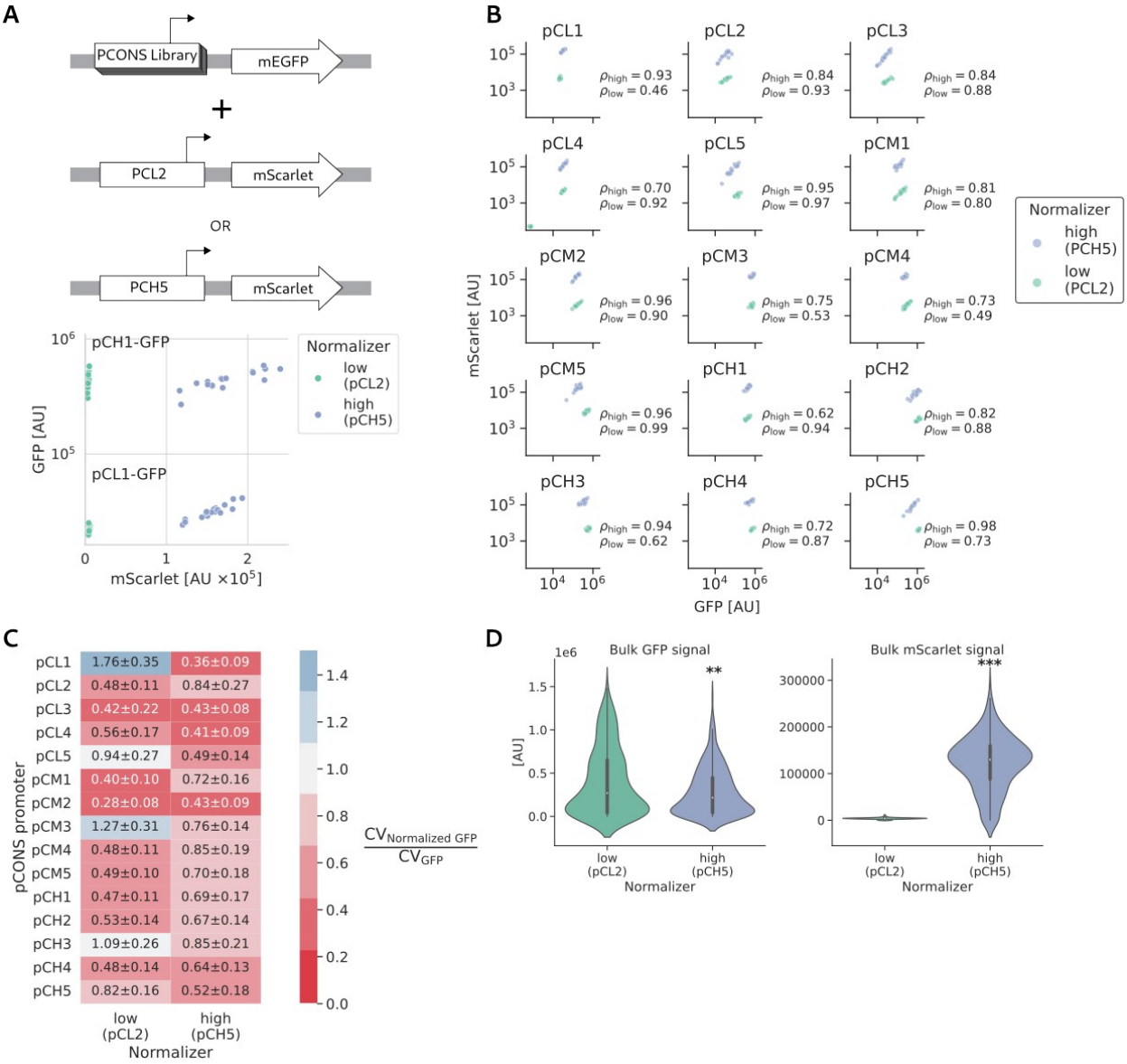

Figure 5. Efficacy of normalization using PCONS and limitation of cellular resources when driving high transgene expression. (A) GFP expression of PCL1 and PCH1 promoters using two different PCONS promoters (PCL2 or PCH5) to normalize GFP signal through mScarlet expression. (B) Correlation of two fluorescent channels when using two (PCL2 or PCH5) different promoters driving mScarlet to normalize the entire PCONS library driving GFP. (C) Bootstrap estimation of the ratio of coefficient of variation for normalized PCONS-driven GFP expression vs unnormalized PCONS-driven GFP expression. (D) Bulk GFP signal for the PCONS library driving GFP when using two different promoters (PCL2 or, PCH5)

for normalization. Significant changes determined through the Kruskal-Wallis test H-test.

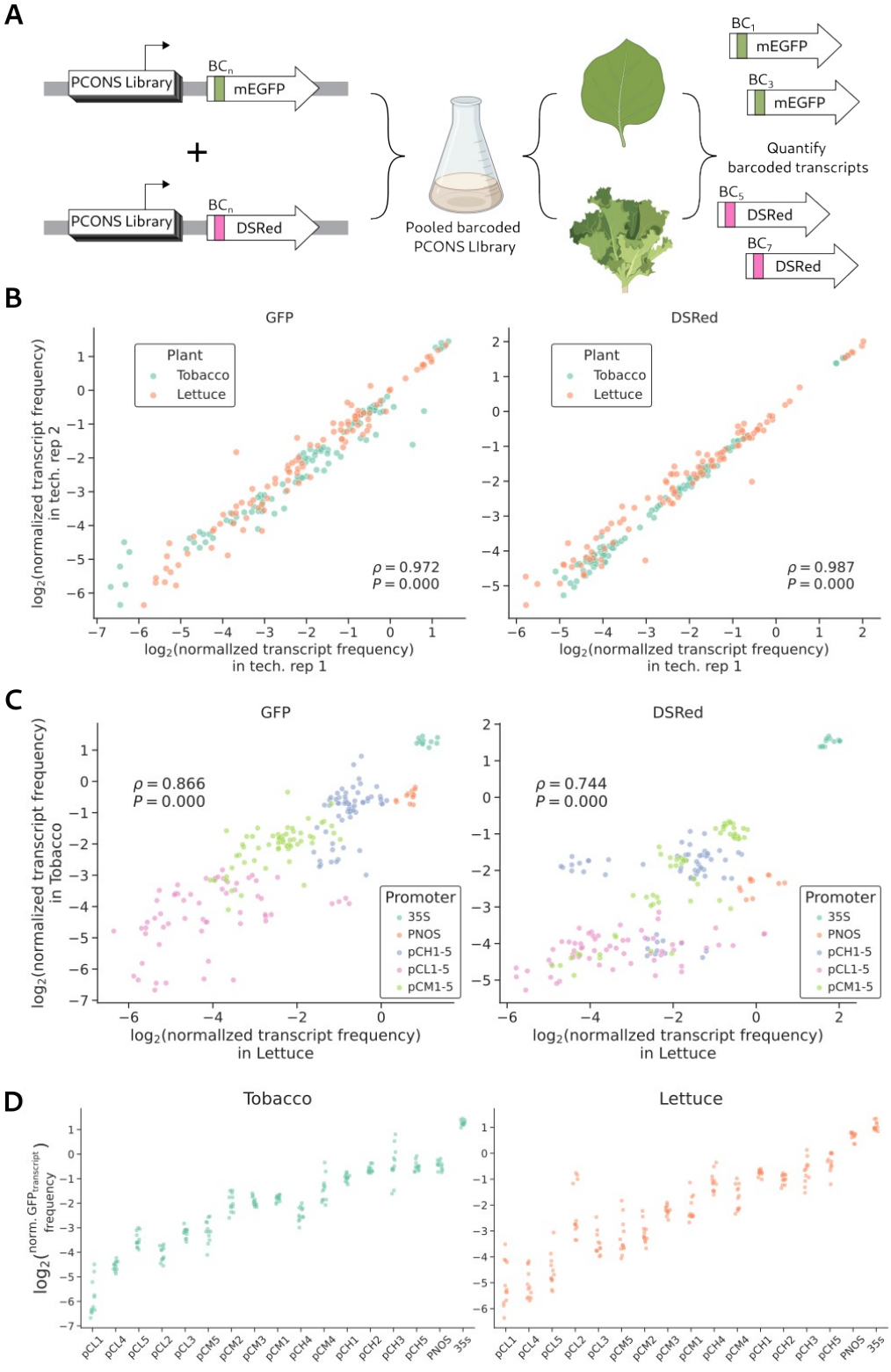

Figure 6. STARR-seq assay reveals conservation of PCONS transcriptional activity between tobacco and lettuce. (A) Design of a barcoded parallel reporter assay to readout PCONS activity driving GFP and DsRed transgene expression in *N. benthamiana* and *L. sativa*. (B) Pearson correlation of technical replicates and shared dynamic range for PCONS driving dual transgenes in *N. benthamiana* and *L. sativa*. Each data point represents the median count of a given PCONS promoter inside a biological replicate and all PCONS promoters are present in the data. (C) Pearson correlation of transgene expression activation by PCONS promoters between *N. benthamiana* and *L. sativa*. Promoters are binned to low (PCL1 to PCL5), medium (PCM1 to PCM5) and high (PCH1 to PCH5). Each data point represents the median count of a given PCONS promoter inside a biological replicate and all PCONS promoters are present in the data. (D) Activation strength of individual PCONS in transcribing GFP. Each data point represents the median count of a given PCONS promoter inside a biological replicate and all PCONS promoters are present in the data.

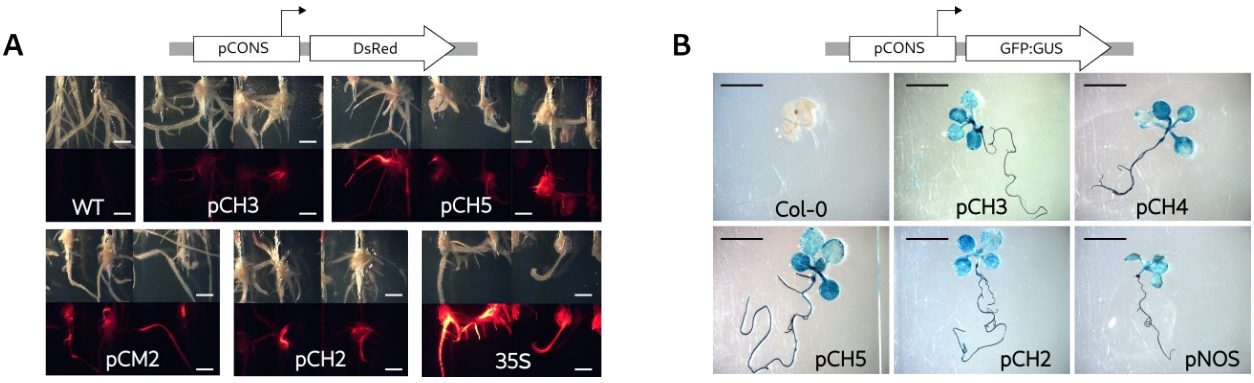

Figure 7. PCONS promoters drive stable transgene expression in Arabidopsis and Medicago (A) Hairy root transformation of Medicago expressing DsRed driven with various PCONS promoters. White bars indicate 5mm. (B) GUS staining of T2 Arabidopsis seedlings with stable integrated GFP:GUS expression cassettes driven with various PCONS promoters. Black bars indicate 1cm.

Tables:

Table 1. PCONS library TAIR IDs and putative gene functions

PCONS ID	TAIR ID	Gene Function
pCH1	AT1G23490	ADP-Ribosylation factor 1A
pCH2	AT1G07890	Ascorbate peroxidase 1
pCH3	AT4G13930	Serine Hydroxymethyltransferase 4
pCH4	AT1G15690	AT V-PPase 3
pCH5	AT5G11740	Arabinogalactan protein 15
pCM1	AT2G44060	LATE EMBRYOGENESIS ABUNDANT 26
pCM2	AT4G40040	HTR5 histone
pCM3	AT3G15730	PHOSPHOLIPASE D ALPHA 1
pCM4	AT3G13920	eukaryotic translation initiation factor 4A-1
pCM5	AT3G55440	CYTOSOLIC ISOFORM TRIOSE PHOSPHATE ISOMERASE
pCL1	AT1G76200	NADH dehydrogenase ubiquinone 1 beta subcomplex unit
pCL2	AT2G33040	Gamma subunit of Mt ATP synthase
pCL3	AT4G34110	ARABIDOPSIS POLY(A) BINDING 2
pCL4	AT1G22300	GENERAL REGULATORY FACTOR 10
pCL5	AT3G44110	DNAJ HOMOLOGUE 3