Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas

Fahim Farzadfar,†‡§ Samuel D. Perli,†‡ and Timothy K. Lu*,†‡§∥

†Department of Electrical Engineering & Computer Science and Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States
‡MIT Synthetic Biology Center, 500 Technology Square, Cambridge, Massachusetts 02139, United States
§MIT Microbiology Program, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States
∥Biophysics Program, Harvard University, Cambridge, Massachusetts 02139, United States

Supporting Information

ABSTRACT: Transcriptional regulation is central to the complex behavior of natural biological systems and synthetic gene circuits. Platforms for the scalable, tunable, and simple modulation of transcription would enable new abilities to study natural systems and implement artificial capabilities in living cells. Previous approaches to synthetic transcriptional regulation have relied on engineering DNA-binding proteins, which necessitate multistep processes for construction and optimization of function. Here, we show that the CRISPR/Cas system of Streptococcus pyogenes can be programmed to direct both activation and repression to natural and artificial eukaryotic promoters through the simple engineering of guide RNAs with base-pairing complementarity to target DNA sites. We demonstrate that the activity of CRISPR-based transcription factors (crisprTFs) can be tuned by directing multiple crisprTFs to different positions in natural promoters and by arraying multiple crisprTF-binding sites in the context of synthetic promoters in yeast and human cells. Furthermore, externally controllable regulatory modules can be engineered by layering gRNAs with small molecule-responsive proteins. Additionally, single nucleotide substitutions within promoters are sufficient to render them orthogonal with respect to the same gRNA-guided crisprTF. We envision that CRISPR-based eukaryotic gene regulation will enable the facile construction of scalable synthetic gene circuits and open up new approaches for mapping natural gene networks and their effects on complex cellular phenotypes.

KEYWORDS: synthetic transcription factors, CRISPR/Cas9, RNA-guided multiplex gene regulation, synthetic gene regulation, endogenous gene regulation, synthetic biology

Complex and sophisticated phenotypes in eukaryotic cells manifest from layered regulatory networks and specific expression programs involving the regulated transcription of many genes.1 As major players in these networks, eukaryotic transcriptional factors (TFs) can integrate multiple signals and perform complex, combinatorial functions on promoters, where regulatory information is encoded in the form of binding sites for TFs and interactions between TFs, to modulate gene expression patterns.1–3

Rewiring endogenous transcriptional networks by natural or synthetic TFs is a powerful strategy for interrogating cellular functions and controlling cellular phenotypes.4–14 Previously, natural DNA-binding domains (DBDs, mainly from bacterial sources, such as TetR, LacI, and LexA) have been used to recruit effector (e.g., activator and repressor) domains to the regulatory regions of eukaryotic genes in order to modulate their transcription.15–17 This necessitates the placement of DBD-specific operator site(s) in the cis-regulatory region of the promoters for specific genes which is a labor- and time-intensive process, especially if the regulation of multiple genes is desired. Moreover, engineering and modulating complex transcriptional networks requires tunable, extensible, and orthogonal transcription factors. However, only a few orthogonal variants of natural DBDs are well-characterized and changing their specificity has proven to be challenging.18 As such, the use of natural DBD-based TFs for wiring complex transcriptional networks and synthetic gene circuits has been limited.

To address these limitations, synthetic TFs based on Zinc Fingers (ZF s) and Transcriptional Activator-Like Effectors (TALEs) have been developed.19–25 The ability to program the specificity of ZFs and TALEs to potentially target any sequence makes these DBDs appealing for designing libraries of orthogonal transcription factors. Synthetic ZF- and TALE-based TFs have been shown to work in a wide range of
However, obtaining a TF for a given target site requires tedious selection processes or multistage DNA assembly protocols. Furthermore, the scale of regulation that can be achieved by these TFs is potentially limited by the metabolic burden imposed on the cells and the number of TFs that can be simultaneously encoded in a given cell.

Here, we present a strategy for modulating eukaryotic transcription at natural and synthetic promoters using programmable and tunable synthetic transcription factors based on a bacterial CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas system. Many bacteria use CRISPR-based immune systems to degrade genetic materials of invading phages. CRISPR-based systems can be used for genetic editing and regulation in eukaryotes; however, obtaining a TF for a given target site requires tedious selection processes or multistage DNA assembly protocols. Furthermore, the scale of regulation that can be achieved by these TFs is potentially limited by the metabolic burden imposed on the cells and the number of TFs that can be simultaneously encoded in a given cell.

Here, we present a strategy for modulating eukaryotic transcription at natural and synthetic promoters using programmable and tunable synthetic transcription factors based on a bacterial CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas system. Many bacteria use CRISPR-based immune systems to degrade genetic materials of invading phages. In these systems, short RNAs expressed from CRISPR loci are used to target an endonuclease protein (Cas9) against invading genetic material. Recently, it has been shown that Cas9 can be used as a programmable tool for genome editing across various organisms. In this context, small customizable guide RNAs (gRNAs) can be used to program and target Cas9 endonuclease to specific loci in living cells to induce double (or single)-stranded breaks in DNA. Upon cleavage, error-prone or template-directed repair pathways are triggered, generating variants of the original target loci. Qi et al. showed that an endonuclease-deficient Cas9 (dCas9, with D10A H841A mutations relative to the wild-type Cas9) can be used as a programmable “CRISPRi” tool for gene silencing in Escherichia coli. When targeted to a promoter or ORF of a gene of interest, dCas9 can block progression of RNA polymerase and hence silence expression of the targeted gene. They also provide evidence that CRISPRi is functional in human cells, albeit with much lower efficiency compared with E. coli. In addition, Bikard et al. demonstrated that along with programmed transcriptional repression, transcriptional activation can be achieved in E. coli by fusing the omega subunit of RNA polymerase to the endonuclease-deficient Cas9.

Here, we achieved versatile, programmable, and multiplexable tools for gene regulation in eukaryotes by functionalizing CRISPR/Cas-based systems for eukaryotic transcriptional regulation in S. cerevisiae. (A) CrisprTF (dCas9_VP64) expression is induced by growing cells in galactose (Gal) + anhydrotetracycline (aTc) media. CrisprTFs are guided to the target sites by guide RNAs (gRNAs), which are constitutively expressed from the pRPR1 promoter and bind to the respective target sites. Specificity of crisprTFs is primarily determined by the 20 bp Specificity Determinant Sequence (SDS) at the 5′-end of the gRNA along with the presence of a Proto-spacer Adjacent Motif (PAM) (NGG) at the target site. Map of pCYC1m illustrating the relative positions of known regulatory elements. TATA: TATA box. TSS: Transcription Start Site. KS: Kozak Sequence. Blue lines indicate target sites for each gRNA (c1-c8). (C) Left panel: Regulation of gfp expression from pCYC1m by crisprTFs based on the individual gRNAs shown in (B). Yeast cells expressing crisprTFs and containing the reporter construct were transformed with plasmids expressing gRNAs labeled as shown in the x-axis. Targeting crisprTFs to sequences upstream of the TATA boxes (by c3, c4, and c6 gRNAs) resulted in higher gfp expression than the no gRNA control. On the other hand, targeting crisprTFs to sequences spanning the TATA box and the Kozak sequence (by c1, c6, and c7 gRNAs) resulted in reduced gfp expression relative to the no gRNA control. Error bars indicate the standard error of the mean for three independent biological replicates. Asterisks (*) on each bar indicate statistically significant changes in gfp expression relative to the no gRNA control (based on the two-sided Welch’s t test, p-value < 0.05). Right panel: Coexpression of multiple gRNAs resulted in synergistic gene regulation. Pairwise combinations of non-neutral gRNAs were expressed from pRPR1 promoters on pRS423 and pRS425 backbones. Green and red asterisks (*) indicate statistically significant changes in gfp expression in samples with coexpressed gRNAs relative to the first gRNA only and the second gRNA only, respectively (two-sided Welch’s t test, p-value < 0.05).
ing dCas9 with effector domains and targeting both natural and synthetic promoters. As a proof of concept, we made an RNA-guidable transcription factor by fusing dCas9 to an activator domain. Using this CRISPR-based transcription factor (crisprTF), we teased apart the regulatory maps of several natural eukaryotic promoters (in *Saccharomyces cerevisiae* and HEK293T cells) without the need to modify promoter architectures. Unlike previous generations of customizable DBDs (i.e., ZFs and TALEs) that require multistage design and cloning strategies, crisprTFs can be readily customized and retargeted to different loci and regulatory regions in vivo using specific gRNAs with homology to target sites (Figure 1). dCas9 thus offers a powerful tool for targeting functions of interest to specific genomic loci in living cells, which can potentially be used to regulate gene expression at will, construct scalable synthetic gene circuits, or rewire endogenous regulatory networks.

To implement crisprTFs in *Saccharomyces cerevisiae*, we fused the SV40 nuclear localization sequence (NLS) and four tandem copies of Herpes Simplex Viral Protein 16 (VP64, a commonly used eukaryotic transcription activator domain) to a codon-optimized *S. pyogenes* dCas9 (Figure 1A).21 The crisprTF cassette was then cloned under the control of pTPGI, a synthetic promoter which can be induced by growing cells in galactose + anhydrotetracycline (aTc) media23 and integrated into the yeast genome. To assess the activity of crisprTF, *gfp* was placed under the control of a minimal CYC1 promoter (pCYC1m) and the whole cassette was integrated into the yeast genome. pCYC1m retains one of the two endogenous TATA boxes of the wild-type CYC1 promoter and lacks binding sites for endogenous regulatory factors in the upstream activating sequence (UAS).42,43 gRNAs were expressed constitutively from the RNA polymerase III-dependent pRPR1 promoter and the 3′-ends of the gRNAs were defined by the pRPR1 terminator.44

The expression of gRNAs targeting different regions in the pCYC1m (as shown in Figure 1B) resulted in various statistically significant levels of reporter fluorescence compared to the no gRNA control (Figure 1C, left panel). Targeting crisprTFs to sequences upstream of the TATA box (by m1, m2, m6, and m7 gRNAs) resulted in higher *gfp* expression compared with the no gRNA control. Error bars indicate the standard error of the mean for three independent biological replicates. Asterisks (*) on each bar indicate statistically significant changes in *gfp* expression relative to the no gRNA control (based on the two-sided Welch’s t test, p-value <0.05).

Figure 2. Regulation of *yfp* expression from a minimal MLP promoter (pMLPm) by crisprTFs in HEK293T cells. (A) dCas9_VP64 is expressed in HEK293T cells by the pCMV promoter and directed to target sequences in pMLPm. The mKATE (red) and mBFP2 (blue) fluorophores act as flow-cytometry gating controls for successful plasmid transfections. (B) Map of pMLPm illustrating the relative positions of known regulatory elements. Blue lines indicate target sites for each gRNA. (C) Regulation of *yfp* expression from pMLPm by crisprTFs based on the gRNAs shown in (B). HEK293T cells were cotransfected with the plasmids shown in (A), with specific gRNAs labeled as shown in the x-axis. Targeting crisprTFs to sequences upstream of the TATA box (by m1, m2, m6, and m7 gRNAs) resulted in higher *yfp* expression compared with the no gRNA control. Error bars indicate the standard error of the mean for three independent biological replicates. Asterisks (*) on each bar indicate statistically significant changes in *yfp* expression relative to the no gRNA control (based on the two-sided Welch’s t test, p-value <0.05).
the TATA box and the transcription start site (using c2 and c6 gRNAs), likely due to interference of crisprTFs with the formation of the transcriptional initiation complex. No activation was observed with any of the eight tested gRNAs when dCas9, without a fused activator domain, was targeted to pCYC1m (Figure S1, Supporting Information). All the tested gRNAs in this strain repressed *gfp* expression to some extent (Figure S1, Supporting Information), with highest repression observed with c6 and c7. These results demonstrate that dCas9 is able to repress transcription but requires an activation domain (VP64) to activate transcription of a target locus and further supports the hypothesis that dCas9 (or as a fusion to VP64) can act as a repressor by interfering with the formation of the transcriptional initiation complex. Similar results were achieved with the GAL1 promoter (pGAL1) and its variants, where targeting crisprTFs to sequences upstream and downstream of TATA box led to activation and repression of the GFP reporter, respectively (Figure S2, Supporting Information). These results indicate that a single crisprTF can be programmed to act as both an activator and a repressor by targeting it to different positions across endogenous promoters.

The activity of a promoter is determined by combinatorial interactions between transcriptional regulatory factors bound to that promoter. We thus investigated the effects of binding of multiple crisprTFs targeted to the same promoter. To this end, pairwise combinations of gRNAs with non-neutral effects (those that showed either activation or repression in the left panel of Figure 1C) were coexpressed. As shown in the right panel of Figure 1C, coexpression of repressor gRNAs resulted in synergistic repression of the reporter (up to 7× repression was achieved with coexpression of the c5 and c6 pair as well as the c6 and c7 pair). On the other hand, when a repressor gRNA was coexpressed with an activator gRNA (e.g., the c3 and c6 pair), an intermediate level of GFP expression was achieved, indicating an antagonistic interaction between the two gRNAs. Moreover, the effects of repressor gRNAs were dominant over activator gRNAs, suggesting that interruption of the formation of the transcription initiation complex has a stronger effect than activation. Coexpression of two activator gRNAs (e.g., c3 and c4) did not result in synergistic activation of the reporter, which suggests that the relative positions and interactions of bound activators are important for determining synergistic activation. Consistent with our results, it has been shown that synergistic activation from synthetic promoters with multiple GAL4 operator sites depends on the distance and helical phase of the operator sites. Furthermore, in another study it has been shown that not all of the combinations of TALE-activators resulted in synergistic repression of the reporter (up to 70-fold activation with 12 gRNAs or gRNAs targeting the CCAAT box or the GC-box gRNA). However, in our study we found that the repression of pPGK1 promoter was not significant when gRNAs targeting the CCAAT box or the GC-box were used.

We next sought to investigate the activity of crisprTFs in human cells. To this end, a human-codon-optimized crisprTF cassette was placed on a plasmid under the control of the constitutive cytomegalovirus immediate-early promoter (pCMV). The gRNAs were expressed constitutively from a separate plasmid by the RNA polymerase III-dependent U6 promoter (pU6), as previously described. After transfection of these plasmids into HEK293T cells, we investigated the regulatory architecture of the minimal adenovirus major late promoter (pMLPm) by targeting crisprTFs to different positions across this promoter (Figure 3A). YFP was used as the readout for pMLPm promoter activity.

Figure 3. CrispTF-mediated repression of the constitutive pPGK1 promoter in HEK293T cells. (A) Map of the pPGK1 promoter illustrating the relative positions of known regulatory elements. (B) CrispTF-based targeted repression of the constitutive pPGK1 promoter. Constructs expressing different dCas9-based proteins (dCas9, dCas9-VP64, and dCas9-KRAB) were cotransfected with plasmids containing pPGK1_mKATE and constructs expressing no gRNAs or gRNAs targeting the CCAAT box or the GC-box gRNA. Significant repression of the pPGK1 promoter relative to the no gRNA control was observed with all of the three different dCas9 constructs (dCas9, dCas9-VP64, and dCas9-KRAB). Error bars indicate the standard error of the mean for three independent biological replicates.
Synergistic and tunable activation of synthetic promoters with arrayed operator sites upstream of pCYC1m in S. cerevisiae and pMLPm in HEK293T cells using crisprTFs. (A) A schematic view of the pCYC1m synthetic promoter with three a1_gRNA operator sites (3x DNA sequences recognized by the a1_gRNA) arrayed upstream of pCYC1m, thus named 3x(a1_op)_pCYC1m. (B) Increasing the number of arrayed a1_gRNA operator sites upstream of pCYC1m resulted in higher bfp expression in S. cerevisiae cells expressing the a1_gRNA compared to the no gRNA controls. Error bars indicate the standard error of the mean for three independent biological replicates. (C) A schematic view of the pMLPm synthetic promoter with three a1_gRNA operator sites arrayed upstream of pMLPm, thus named 3x(a1_op)_pMLPm. (D) Increasing the number of arrayed a1_gRNA operator sites upstream of pMLPm resulted in higher yfp expression in HEK293T cells when cotransfected with a1_gRNA and dCas9-Vp64 versus when cotransfected with a1_gRNA and dCas9. Error bars indicate the standard error of the mean for three independent biological replicates.

Figure 4. Synergistic and tunable activation of synthetic promoters with arrayed operator sites upstream of pCYC1m in S. cerevisiae and pMLPm in HEK293T cells using crisprTFs. (A) A schematic view of the pCYC1m synthetic promoter with three a1_gRNA operator sites (3x DNA sequences recognized by the a1_gRNA) arrayed upstream of pCYC1m, thus named 3x(a1_op)_pCYC1m. (B) Increasing the number of arrayed a1_gRNA operator sites upstream of pCYC1m resulted in higher bfp expression in S. cerevisiae cells expressing the a1_gRNA compared to the no gRNA controls. Error bars indicate the standard error of the mean for three independent biological replicates. (C) A schematic view of the pMLPm synthetic promoter with three a1_gRNA operator sites arrayed upstream of pMLPm, thus named 3x(a1_op)_pMLPm. (D) Increasing the number of arrayed a1_gRNA operator sites upstream of pMLPm resulted in higher yfp expression in HEK293T cells when cotransfected with a1_gRNA and dCas9-Vp64 versus when cotransfected with a1_gRNA and dCas9. Error bars indicate the standard error of the mean for three independent biological replicates.

Synergistic and tunable activation of synthetic promoters with arrayed operator sites upstream of pCYC1m in S. cerevisiae and pMLPm in HEK293T cells using crisprTFs. (A) A schematic view of the pCYC1m synthetic promoter with three a1_gRNA operator sites (3x DNA sequences recognized by the a1_gRNA) arrayed upstream of pCYC1m, thus named 3x(a1_op)_pCYC1m. (B) Increasing the number of arrayed a1_gRNA operator sites upstream of pCYC1m resulted in higher bfp expression in S. cerevisiae cells expressing the a1_gRNA compared to the no gRNA controls. Error bars indicate the standard error of the mean for three independent biological replicates. (C) A schematic view of the pMLPm synthetic promoter with three a1_gRNA operator sites arrayed upstream of pMLPm, thus named 3x(a1_op)_pMLPm. (D) Increasing the number of arrayed a1_gRNA operator sites upstream of pMLPm resulted in higher yfp expression in HEK293T cells when cotransfected with a1_gRNA and dCas9-Vp64 versus when cotransfected with a1_gRNA and dCas9. Error bars indicate the standard error of the mean for three independent biological replicates.

Synergistic and tunable activation of synthetic promoters with arrayed operator sites upstream of pCYC1m in S. cerevisiae and pMLPm in HEK293T cells using crisprTFs. (A) A schematic view of the pCYC1m synthetic promoter with three a1_gRNA operator sites (3x DNA sequences recognized by the a1_gRNA) arrayed upstream of pCYC1m, thus named 3x(a1_op)_pCYC1m. (B) Increasing the number of arrayed a1_gRNA operator sites upstream of pCYC1m resulted in higher bfp expression in S. cerevisiae cells expressing the a1_gRNA compared to the no gRNA controls. Error bars indicate the standard error of the mean for three independent biological replicates. (C) A schematic view of the pMLPm synthetic promoter with three a1_gRNA operator sites arrayed upstream of pMLPm, thus named 3x(a1_op)_pMLPm. (D) Increasing the number of arrayed a1_gRNA operator sites upstream of pMLPm resulted in higher yfp expression in HEK293T cells when cotransfected with a1_gRNA and dCas9-Vp64 versus when cotransfected with a1_gRNA and dCas9. Error bars indicate the standard error of the mean for three independent biological replicates.

Synergistic and tunable activation of synthetic promoters with arrayed operator sites upstream of pCYC1m in S. cerevisiae and pMLPm in HEK293T cells using crisprTFs. (A) A schematic view of the pCYC1m synthetic promoter with three a1_gRNA operator sites (3x DNA sequences recognized by the a1_gRNA) arrayed upstream of pCYC1m, thus named 3x(a1_op)_pCYC1m. (B) Increasing the number of arrayed a1_gRNA operator sites upstream of pCYC1m resulted in higher bfp expression in S. cerevisiae cells expressing the a1_gRNA compared to the no gRNA controls. Error bars indicate the standard error of the mean for three independent biological replicates. (C) A schematic view of the pMLPm synthetic promoter with three a1_gRNA operator sites arrayed upstream of pMLPm, thus named 3x(a1_op)_pMLPm. (D) Increasing the number of arrayed a1_gRNA operator sites upstream of pMLPm resulted in higher yfp expression in HEK293T cells when cotransfected with a1_gRNA and dCas9-Vp64 versus when cotransfected with a1_gRNA and dCas9. Error bars indicate the standard error of the mean for three independent biological replicates.

Synergistic and tunable activation of synthetic promoters with arrayed operator sites upstream of pCYC1m in S. cerevisiae and pMLPm in HEK293T cells using crisprTFs. (A) A schematic view of the pCYC1m synthetic promoter with three a1_gRNA operator sites (3x DNA sequences recognized by the a1_gRNA) arrayed upstream of pCYC1m, thus named 3x(a1_op)_pCYC1m. (B) Increasing the number of arrayed a1_gRNA operator sites upstream of pCYC1m resulted in higher bfp expression in S. cerevisiae cells expressing the a1_gRNA compared to the no gRNA controls. Error bars indicate the standard error of the mean for three independent biological replicates. (C) A schematic view of the pMLPm synthetic promoter with three a1_gRNA operator sites arrayed upstream of pMLPm, thus named 3x(a1_op)_pMLPm. (D) Increasing the number of arrayed a1_gRNA operator sites upstream of pMLPm resulted in higher yfp expression in HEK293T cells when cotransfected with a1_gRNA and dCas9-Vp64 versus when cotransfected with a1_gRNA and dCas9. Error bars indicate the standard error of the mean for three independent biological replicates.

Synergistic and tunable activation of synthetic promoters with arrayed operator sites upstream of pCYC1m in S. cerevisiae and pMLPm in HEK293T cells using crisprTFs. (A) A schematic view of the pCYC1m synthetic promoter with three a1_gRNA operator sites (3x DNA sequences recognized by the a1_gRNA) arrayed upstream of pCYC1m, thus named 3x(a1_op)_pCYC1m. (B) Increasing the number of arrayed a1_gRNA operator sites upstream of pCYC1m resulted in higher bfp expression in S. cerevisiae cells expressing the a1_gRNA compared to the no gRNA controls. Error bars indicate the standard error of the mean for three independent biological replicates. (C) A schematic view of the pMLPm synthetic promoter with three a1_gRNA operator sites arrayed upstream of pMLPm, thus named 3x(a1_op)_pMLPm. (D) Increasing the number of arrayed a1_gRNA operator sites upstream of pMLPm resulted in higher yfp expression in HEK293T cells when cotransfected with a1_gRNA and d Cas9-Vp64 versus when cotransfected with a1_gRNA and dCas9. Error bars indicate the standard error of the mean for three independent biological replicates.
Figure 5. Inducible crisprTF-guided activation of synthetic promoters. (A) Schematic of the αTc-inducible pRPR1_TetO promoter. Expression of dCas9_VP64 is driven by the galactose-inducible pGAL1 promoter. A TetR operator site (1xTetO) was placed in the pRPR1 promoter to make an αTc-responsive pRPR1_TetO promoter. Addition of αTc releases TetR-mediated repression on the pRPR1_TetO promoter and results in a1_gRNA expression. (B) αTc-dependent bfp expression from a synthetic 6x(a1_op)_pCYC1m promoter. S. cerevisiae cells containing the circuit shown in (A) were grown in galactose media with either 250 ng/mL αTc or no αTc. Error bars indicate the standard error of mean for three biological replicates.

repression functions can be achieved with only one transcription factor, by targeting dCas9_VP64 fusions to different regulatory sequences along a promoter. Our finding that one can activate or repress the expression of a gene of interest by directing a single protein to different positions of a promoter is advantageous for the efficient design of synthetic transcriptional networks or rewiring natural ones. This property obviates the need for using separate orthogonal Cas9 protein fusions as activators and repressors. Furthermore, we demonstrate that more sophisticated regulatory motifs, such as small-molecule responsive modules can be built for crisprTFs, thus enabling external control of crisprTF-based transcriptional circuits. Such synthetic modules can be interfaced with other regulatory elements to achieve more complex regulation for synthetic biology.

Our results show that dCas9 can be used as a customizable RNA-guided DNA-binding platform for the regulation of gene expression at natural and synthetic promoters in eukaryotic cells. The ease of design and expression of customized gRNAs in comparison to ZFs and TALEs make CRISPR-based transcription factors appealing as synthetic TFs for modulating endogenous gene expression as well as for synthetic biology. The ability to customize the target site of dCas9 via the expression of short gRNAs obviates the need to engineer multiple orthogonal DBDs in order to construct complex transcriptional circuits. This could potentially reduce the overall metabolic burden on cells and enable the integration of more complex synthetic computation and logic within living cells.

More complex regulatory and logic circuits, such as cascades and complex digital logic gates can be built by layering crisprTFs. The possibility of integrating multiple inputs at a single promoter expands the regulatory potential and provides us with increased flexibility that can be leveraged while designing synthetic transcriptional networks or rewiring endogenous pathways.

Furthermore, since both activation and repression functions can be achieved with crisprTFs, the crisprTF platform may be advantageous compared to noncoding RNA-based gene regulatory platforms where only repression can be achieved. In a way, crisprTFs combine the multiplexability of RNA-based regulatory approaches with the flexibility and rich functionality repertoire of protein-based gene regulatory approaches: Cas9 can be functionalized with regulatory domains of interest (e.g., activation, repression, or epigenetic effector) and then be targeted to multiple loci using different gRNAs.

Future work is needed to define the range of effector domains that can be used with dCas9 for a variety of regulatory functions, including transcriptional regulation and epigenetic modifications. In addition, the identification, characterization, and optimization of Cas9 homologues or evolved variants may enable enhanced activity and specificity of this system. Moreover, the ability to synthesize random libraries of gRNAs opens the possibility for high-throughput perturbations of transcriptional networks and screening for desirable phenotypes. Ultimately, we envision that crisprTFs will enable the regulation and perturbation of natural transcriptional networks as well as the construction of complex synthetic circuits at an unprecedented speed and scale.

■ METHODS

Strain and Plasmid Construction. Saccharomyces cerevisiae. dCas9 (endonuclease-deficient Cas9, with D10A and H841A mutations relative to the wild-type sequence of S. pyogenes Cas939) with an N-terminal SV40 nuclear localization signal (NLS) was codon-optimized for expression in S. cerevisiae and cloned into a pRS314 backbone under control of the pTPGI promoter.23 The RNA-guided transcription factors (crisprTFs) were built by fusing four repeats of the minimal domain of the herpes simplex viral protein 16 (VP16) to the C-terminus of dCas9 (dCas9 VP64). The crisprTF-expressing plasmid was then integrated into the TRP1 locus of S. cerevisiae W303.

The reporter plasmids were built by cloning yeast-enhanced gfp under the control of the wild-type or modified pCYC1m promoter into pRS406 using one-step Gibson assembly. The reporters for the multiple-gRNA-binding-site experiment (Figure 4A) were built by cloning the corresponding number of binding sites upstream of the pCYC1m promoter driving production of EBFP2. All reporters were integrated into the BL1 locus of the integrated crisprTF plasmid.

To build gRNA-expressing plasmids, empty gRNA expressing vectors were first made by cloning the pRPR1 promoter (an RNA-polymerase-III-dependent promoter34), the gRNA handle (flanked by HindIII and XhoI sites), and the RPR terminator

ACS Synthetic Biology
Figure 6. Constructing orthogonal crisprTF-responsive promoters. (A) A schematic view of gRNAs targeting the wild-type and modified pCYC1m promoters in yeast. Only the c1-c8 gRNAs have perfect homology to the target sequences in pCYC1m. On the other hand, c1, c4, c5, and cm1-cm6 gRNAs have perfect homology to the sequences in the pCYC1m(modified) promoter. Mismatches between pCYC1m and pCYC1m(modified) are marked by asterisks (*). (B) pCYC1m only responds to gRNAs that are perfectly matching gRNAs (c1-c8 gRNAs) and not to those that contain mismatches (cm1-cm6 gRNAs). The pCYC1m(modified) promoter responds to the cm1-cm6 gRNAs. Those gRNAs that bind to the sequences upstream of the TATA boxes activate gfp expression and those that target sequences downstream of the TATA boxes repress gfp expression. The c1 and c5 gRNAs have similarly neutral effects on both the wild-type and modified promoters. Error bars indicate standard error of the mean for three independent biological replicates. Asterisks (*) on each bar indicate statistically significant activation or repression relative to no gRNA controls (based on the two-sided Welch’s t test, p-value < 0.05). (C) Heat map illustrating the orthogonality of crisprTFs in human cells. Plasmids encoding three orthogonal gRNAs (a1, a2, and a3 gRNAs) were cotransfected into HEK293T cells along with one of the three reporter plasmids (each encoding 4x operator sites for a given gRNA) upstream of pMLPm promoter driving yfp expression. Only cognate interactions between gRNAs and target binding sites resulted in significant activation. The standard error of the mean of YFP fluorescence for three independent biological replicates is indicated in each cell of the heat-map plot.
into the Sacl and KpnI sites of either the pRS423 or pRS425 plasmid using one-step Gibson assembly. The specificity determinant sequence (SDS) for each gRNA was then cloned into the HindIII site of these vectors by one-step Gibson assembly. Sequences of the constructs used in this study are listed in Table S1, Supporting Information.

HEK293T Cells. To construct the mammalian dCas9_VP64 expressing plasmid, we first introduced D10A and H841A mutations into hCas9 (Addgene, Plasmid #41815). Then, three repeats of SV40 NLS (3xNLS) were fused to the C-terminus of the mutated hCas9 using a PCR-based assembly protocol. Using a multipart Gibson assembly protocol, the immediate-early promoter of cytomegalovirus (pCMV), dCas9_3xNLS, VP64, and SV40 polyA terminator were cloned into the NotI site of the pGS-Luc plasmid (Promega). To monitor successfully transfected cells by flow cytometry, we replaced the original luciferase gene in pGS-Luc with mKATE (Evrogen). The resulting pPGK1_mKATE cassette served as a constitutive fluorescent protein control that was used to gate for the presence of the crisprTF-expressing plasmid with flow cytometry.

The gRNA expression plasmids were constructed by cloning the 138 bp human U6 promoter (an RNA-polymerase-III-dependent promoter48), along with the gRNA handle and terminator into a plasmid containing pPGK1-cBFP2 flanked by the SV40 polyA terminator (a gift from Lior Nissim). A Sacl site was placed at the 3′-end of the U6 promoter to enable the cloning of different specificity determining sequences for each gRNA. The reporters were assembled into the gRNA-expressing plasmid through a one-step Gibson assembly reaction, where the upstream polyadenylation signal and transcriptional pause site from pG5-Luc, along with a 41 bp, minimal adenovirus type 2 major late promoter (pMLPm), mYFP, and HSV polyA signal were cloned into the AatII site of the gRNA-expressing plasmids.

For the synthetic promoter experiments, additional gRNA operator sites were cloned in the Nhel site upstream of the pMLPm promoter (see Supporting Information). For the repression experiments, dCas9 KRAB was constructed by cloning a 366 bp KRAB domain to the C-terminus of dCas9. GCCACC was used as the Kozak sequence for the expression of dCas9_VP64, mYFP, eBFP2, and mKATE.

Unless directly targeted by gRNAs for repression assays, the mKATE fluorescent protein on the crisprTF-expression plasmid and the eBFP2 fluorescent protein on the reporter/gRNA plasmid served as our gating controls for flow cytometry analysis.

Fluorescence Assays. To assess expression of the reporter constructs, yeast cells expressing different gRNAs (or no gRNA as control) were grown overnight (900 rpm, 30°C) in 96-deep-well plates in yeast minimal media supplemented with glucose with appropriate selection (three independent cultures for each sample). Ten microliters of these cultures were then transferred into fresh media supplemented with galactose +250 ng/mL anhydrotetracycline (tCt) and grown for 20 h (900 rpm, 30°C) before analysis by flow cytometry.

For the human cell culture experiments, HEK293T kidney epithelial cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS), 1% glutamine, and 1% penicillin/streptomycin. Cells were grown under 5% CO2 at 37°C. HEK293T cells were transfected with Fugene-HD transfection reagent (Promega) and assayed for gene expression with flow cytometry at 48 h post transfection.

An LSR Fortessa II flow cytometer equipped with 405 nm, 488 nm, and 561 nm lasers was used for all the experiments. GFP/YFP, BFP, and mKATE levels were detected using 488/FITC, 405/Blue, and 561/TX-red laser/filter sets, respectively. All samples were uniformly gated by forward and side scatter. Additional gating for the presence of red and blue fluorophores was applied to the HEK293T samples to ensure only cells successfully transfected with both the crisprTF and the reporter/gRNA plasmids are analyzed. For each gated sample, the mean fluorescence per cell was calculated. Three independent biological samples were used to calculate the mean and standard error of the mean for each data point.

**REFERENCES**

A synthetic genetic edge detection program.


