Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements

Jason Fontana*,[1], Chen Dong*,[2], Cholpisit Kiattisewee[1], Venkata P. Chavali[3], Benjamin I. Tickman[1], James M. Carothers+,[1,3,4], Jesse G. Zalatan+,[2,3,4]

* these authors contributed equally

+ To whom correspondence should be addressed: zalatan@uw.edu or jcaroth@uw.edu

[1]Molecular Engineering & Sciences Institute

[2]Department of Chemistry

[3]Department of Chemical Engineering

[4]Center for Synthetic Biology

University of Washington, Seattle, WA 98195, USA

Abstract

In bacterial systems, CRISPR-Cas transcriptional activation (CRISPRa) has the potential to dramatically expand our ability to regulate gene expression, but we lack predictive rules for designing effective gRNA target sites. Here, we identify multiple features of bacterial promoters that impose stringent requirements on CRISPRa target sites. Notably, we observe narrow, 2-4 base windows of effective sites with a periodicity corresponding to one helical turn of DNA, spanning ~40 bases and centered ~80 bases upstream of the TSS. However, we also identify two features suggesting the potential for broad scope: CRISPRa is effective at a broad range of σ^{70}-family promoters, and an expanded PAM dCas9 allows the activation of promoters that cannot be activated by *S. pyogenes* dCas9. These results provide a roadmap for future engineering efforts to further expand and generalize the scope of bacterial CRISPRa.

Introduction

Developing tools to activate the expression of arbitrary genes has been transformative for biotechnology and biological research[1]. In metabolic engineering, regulating the timing and levels of the expression of complex multi-gene pathways is critical for reducing cellular burden and improving production of valuable metabolites[2]. To enable these goals, we recently developed a CRISPR-Cas transcriptional activation (CRISPRa) system that is effective in *E. coli*. Our system can be combined with CRISPRi gene repression to programmably target multiple genes for simultaneous activation and repression[3]. While our CRISPRa system can be used with heterologous genes, an outstanding challenge is to understand the rules that define effective target sites at arbitrary promoters in the genome.

To programmably downregulate target genes, we use nuclease defective Cas9 (dCas9) with a guide RNA (gRNA) that specifies a target site on the DNA. Targeting this complex to a promoter or an open reading frame (ORF) results in gene repression (CRISPRi)[4]. To enable simultaneous activation, we use modified guide RNAs, termed scaffold RNAs (scRNAs), that include a 3' MS2 hairpin to recruit a transcriptional activator fused to the MS2 coat protein (MCP)[3]. We can express multiple gRNAs and scRNAs to inhibit and activate genes simultaneously; gRNAs targeted to a promoter or ORF result in CRISPRi and scRNAs targeted to an appropriate site upstream of a minimal promoter result in CRISPRa.

We demonstrate here that the rules for targeting CRISPRa to effective sites in *E. coli* are surprisingly stringent. In prior work, we found that CRISPRa *in E. coli* was effective at target sites located in a narrow 40 base window between 60 and 100 bases upstream of the transcriptional start site (TSS)[3]. Here, we show that multiple factors combine to make the requirements for effective sites even stricter. We demonstrate that the basal promoter strength of the target gene

and the sequence composition between the target site and the minimal promoter can have dramatic effects on gene activation. Further, by scanning the 40 base window at single base resolution, we find sharp peaks of activity and broad regions of inactivity that occur in a periodic 10-11 base pattern, corresponding to one helical turn along the DNA target. The observation that only a few precisely-positioned target sites upstream of the TSS are effective for CRISPRa poses a significant challenge, as many genes will likely lack an NGG PAM sequence at exactly the right position necessary for *S. pyogenes* dCas9. These stringent requirements may explain why CRISPRa and other tools for gene activation in bacteria have lagged far behind comparable tools in eukaryotic systems, where such strict target site requirements are absent[5].

Although the requirements for bacterial CRISPRa target sites pose challenges, our data also demonstrate CRISPRa has the potential to be effective at a broad range of target genes. In addition to σ^{70}–dependent genes, CRISPRa can activate expression from genes that use the σ^{70} family members σ^{38}, σ^{32}, and σ^{24}. We further demonstrate that the strict requirement for a precisely positioned PAM site can be partially overcome using a re-engineered dCas9 protein that targets an expanded set of PAM sequences[6]. Recently, some of the rules that we describe here were independently reported for an alternative bacterial CRISPRa system that can target genes regulated by σ^{54} promoters[7]. Our results demonstrate that this behavior applies to a much broader range of σ^{70}–family promoters, which cover the majority of the E. coli genome[8]. The availability of these complementary systems should further extend the scope of bacterial CRISPRa. More broadly, by systematically defining the rules for effective CRISPRa sites, we identify strategies for improving and generalizing synthetic gene regulation in bacteria.

Results

A SoxS mutant reduces off-target activation

Ideally, a synthetic transcriptional activator should only activate its programmed target genes. The activation domain for our CRISPRa system is SoxS, a native *E. coli* transcription factor that directly binds DNA and activates endogenous gene targets as part of a stress response program[3]. We previously demonstrated that point mutations in the SoxS DNA binding site can reduce activation of endogenous SoxS targets while maintaining CRISPRa activity at a heterologous reporter gene. However, the most effective single point mutants, R93A and S101A, did not completely abolish activity at endogenous targets. To further minimize off-target SoxS activity, we tested a double mutant SoxS(R93A/S101A). This double mutant SoxS retained full CRISPRa activity and showed a reduction in endogenous SoxS-dependent gene expression to levels indistinguishable from background (Figure 1). Thus, SoxS(R93A/S101A) is an effective modular transcriptional effector that can activate gene expression only when recruited to a target gene via the CRISPR-Cas complex.

A distance metric for target sites is not effective

To determine if we could predictably activate endogenous genes with CRISPRa, we selected three candidate genes with appropriately positioned PAM sites upstream of the TSS. Previously, we demonstrated that CRISPRa can activate heterologous promoters up to 50-fold with target sites positioned within a 40 base window between 60 and 100 bases upstream of the transcriptional start site (TSS)[3]. We therefore targeted the CRISPR-Cas complex to the same window upstream of the candidate target genes. First, we targeted the *aroK-aroB* operon, which expresses enzymes involved in aromatic amino acid biosynthesis, whose programmed overexpression could be useful for bioproduction[9]. Targeting the CRISPR-Cas complex to two sites within the optimal 40 base window resulted in no statistically significant increases in gene expression. Further, sites inside and outside of the 40 base window gave similar effects (Figure 2A). Next, we targeted *cysK*, an enzyme involved in cysteine biosynthesis[10]. Similar to what we observed with *aroK-aroB*, targeting three sites within the 40 base window resulted in no

statistically significant increases in gene expression (Figure 2B). Finally, we targeted *ldhA*, an enzyme involved in mixed acid fermentation[11]. We selected 8 sites and observed no apparent relationship between the position of the target site and *ldhA* expression (Supplementary Figure 1). Together, these results suggest that endogenous genes cannot be activated simply by targeting the CRISPR-Cas complex to sites positioned between 60 and 100 bases upstream of the TSS.

There are several possible explanations for our inability to activate endogenous bacterial genes with CRISPRa. First, we originally demonstrated CRISPRa using a relatively weak synthetic promoter. The basal levels of expression of endogenous genes vary significantly[12], and it may be difficult to increase the transcription of genes that are already strongly expressed[13]. In addition, some endogenous target genes might require an alternative sigma factor. Our original reporter gene is controlled by the σ^{70} housekeeping sigma factor, and we do not know if our CRISPRa system is effective at gene targets that use alternative sigma factors. Another possibility is that native transcriptional regulator binding sites near endogenous gene promoters could disrupt CRISPRa. Finally, the optimal distance window metric that we previously identified may have been oversimplified. We initially identified the optimal window from an experiment with target sites spaced 10 bases apart, which may not be sufficient to generalize to any site within the 40 base window. To systematically explore these possibilities, we proceeded to test the efficacy of CRISPRa with a new set of synthetic promoters engineered with variable basal expression levels, alternative sigma factors, variable regulator binding sites, and variable scRNA target site positions.

CRISPRa is sensitive to promoter strength

To evaluate whether the intrinsic strength of the promoter affects CRISPRa, we tested activation on a set of fluorescent reporter genes with minimal promoters spanning a 200-fold

range in basal expression level [http://parts.igem.org] (Figure 3A). We observed the most effective gene activation with a moderately weak J23117 promoter. With the weakest promoters, we could not detect any activation, even though their basal expression levels were only 2-fold weaker than the J23117 promoter. With stronger promoters, we observed progressively smaller CRISPRa-mediated activation of gene expression; the basal expression level increased, while the maximal, CRISPRa-induced expression remained roughly constant. These results indicate that the bacterial CRISPRa activity varies considerably with promoter strength, similar to effects observed in eukaryotic systems[14,15]. Thus, when targeting arbitrary endogenous genes, the level of activation that can be achieved may depend on the basal level of expression of its promoter.

CRISPRa is effective with alternative sigma factors

Bacterial transcription is initiated by a sigma factor binding to the minimal promoter and the RNA polymerase holoenzyme[16].The SoxS activator binds directly to the α subunit of RNA polymerase[17], which suggests that our CRISPRa system could be compatible with genes that are controlled by non-housekeeping sigma factors. To investigate this possibility, we built synthetic promoters regulated by σ^{38} (RpoS), σ^{32} (RpoH), σ^{24} (RpoE), and σ^{54} (RpoN) to compare with our original housekeeping σ^{70} (RpoD) promoter (Figure 3B)[18–21]. CRISPRa was able to activate reporter gene expression when we targeted σ^{38}, σ^{32}, and σ^{24}-dependent promoters; these σ factors are all members of the σ^{70} family. CRISPRa was not active on the σ^{54} promoter, possibly because σ^{54} initiates gene expression using a distinct mechanism that requires additional *cis*-regulatory elements[16]. These results suggest that CRISPRa can activate promoters regulated by non-housekeeping sigma factors such as σ^{38}, σ^{32}, and σ^{24}, and likely other members of the homologous σ^{70} family.

A recent paper described an alternative CRISPRa system that is capable of activating σ^{54}-dependent genes[7], which comprise a small fraction of the genome[8] (Supplementary Figure 2).

The availability of multiple, complementary CRISPRa systems should further extend the scope of bacterial CRISPRa. Both systems effectively activate expression from synthetic and heterologous promoters, and each system has the potential to target a different, non-overlapping set of endogenous genes.

CRISPRa is sensitive to intervening sequence composition

To determine if the sequence composition between the target site and the -35 site affects CRISPRa, we constructed a promoter library with randomized sequences in this intervening region. We analyzed single colonies from this library and observed gene activation with a broad distribution over a 27-fold range (Figure 3C). Although most variant sequences can still be activated (>2-fold) with CRISPRa, the large variation in activity was unexpected because each reporter gene was driven by the same minimal promoter and contained the same scRNA target site. One possible interpretation of this result is that these randomized intervening sequences contain binding sites for endogenous transcriptional regulators; there is evidence that binding sites can emerge with relatively high frequency from random sequences[22]. These sites could potentially affect CRISPRa by directly blocking access to a scRNA target site, by blocking RNA polymerase binding, or by interfering with the ability of a CRISPRa effector protein to engage with RNA polymerase.

To directly test the hypothesis that a bound transcriptional effector can disrupt CRISPRa, we introduced a binding site for the transcriptional repressor TetR upstream of the -35 region[23]. The presence of a bound TetR significantly disrupted CRISPRa-mediated gene activation. Further, adding anhydrotetracycline (aTc), which releases TetR from the DNA, restored CRISPRa activity to the levels observed when TetR was not present (Figure 3D). Because endogenous genes contain binding sites for a variety of transcriptional activators and repressors upstream of

the minimal promoter[24,25], this effect could be contributing to the inconsistent and variable effects we observed when targeting endogenous genes for CRISPRa (Figure 2).

To determine if transcription factor binding sites appear in the library of randomized intervening sequences, we sequenced 29 variants spanning the full range of observed activation levels (Supplementary Table 6). Only five intervening sequences contained exact matches to a known consensus transcription factor binding motif. However, all sequences contained at least one match within a single base of a known motif, and it is well established that DNA binding proteins can recognize sites that deviate from the consesus[26]. There was no significant correlation between gene activation by CRISPRa and the number of these motifs (Spearman rank order correlation r_s = 0.29, p = 0.11, Supplementary Figure 3A), but we note that it is not known which of these motifs actually bind endogenous transcription factors. We did find that intervening sequences that give more effective CRISPRa tend to be more GC-rich, though we do not yet understand the basis for this trend (r_s = 0.42, p = 0.02, Supplementary Figure 3B & C). Nonetheless, these experiments indicate that the composition of the intervening sequence between the CRISPR-Cas complex and the minimal promoter is an important factor determining the level of CRISPRa.

CRISPRa is sharply dependent on single base shifts

Our original hypothesis that optimal target sites are located -60 to -100 bases upstream of the TSS was based on an experiment with scRNA sites spaced every 10 bases[3]. To further test this hypothesis, we targeted the CRISPRa complex to a window from -61 to -113 at single base resolution. We used a reporter gene with 5 scRNA sites located at -61, -71, -81, -91, and -101 relative to the TSS, and we inserted 1-12 bases upstream of the -35 site to generate a set of reporter genes that allowed the CRISPRa complex to target every possible distance in the optimal targeting window. Using this reporter gene set, we found that shifting the target site by 1-3 bases

caused significant decreases in activation (Figure 4A). Shifting the target site further by 4-9 bases decreased expression to levels nearly indistinguishable from background. At 10-11 base shifts, corresponding to one full turn of a DNA helix, gene expression increased again. This periodic positional dependence of CRISPRa extended over the entire -60 to -100 window, with the strongest peaks centered at -81 and -91 and smaller peaks centered at -102 and -70. There is no recovery of activity when the site at -101 is shifted to -111, outside of the -60 to -100 window. This sharp periodic relationship suggests that the criteria for effective target sites are quite stringent, and that both distance and relative periodicity to the TSS are critical factors.

Notably, the distance to the TSS is not the sole determining factor for CRISPRa-mediated expression level. Sites that overlap at the same distance, such as the original -81 site and the -71 site shifted by 10, do not give the same gene expression output (Figure 4A). These discrepancies could arise from intrinsic differences in the activity of the 20 base scRNA target sequence (Supplementary Figure 4) or from the effect of different intervening sequence composition between the scRNA target site and the minimal promoter (Figure 3).

Because we demonstrated that sequence composition can have unexpected effects on CRISPRa (Figure 3), we tested whether the periodicity of CRISPRa was similar in different sequence contexts. We obtained comparable periodic phase dependence when different nucleotide sequences were used to shift the scRNA target site, and when the bases were inserted at a different location in the promoter (Supplementary Figure 5A). Similar results were also obtained when we performed the base shift experiment with a reporter that had a different 5' upstream sequence (Supplementary Figure 5B) or where the minimal BBa_J23117 promoter was replaced by endogenous *aroK* promoter (Supplementary Figure 5C). Further, the sharp positioning dependence was observed when targeting the template or non-template strand of the reporter (Supplementary Figure 5D). Finally, one possible confounding effect could arise if the

basal expression level of the reporter gene changes when bases are inserted, which can affect the efficacy of CRISPRa (Figure 3A). However, we observed that basal expression from the original reporter and the +5 base shifted reporter were indistinguishable (Supplementary Figure 5E). Together, these experiments confirm that bacterial CRISPRa is sensitive to periodicity in multiple different sequence contexts.

In the experiments described above, comparisons between single base shifted scRNA sites were performed with different reporter gene constructs, each with a differing number of inserted bases. To test the positional dependence of CRISPRa at single base resolution in a single reporter construct, we designed an alternative reporter gene with 6 adjacent scRNA target sites between -81 and -86. We again observed sharp drops in gene expression when targeting sites one or more bases away from the optimal site at -81 (Supplementary Figure 5F).

The finding that CRISPRa displays the same ~10 base periodicity as the DNA helix suggests that the angular phase of the CRISPRa complex relative to the minimal promoter is critical for effective activation. Our bacterial CRISPRa system requires a direct interaction between the SoxS activation domain and RNA polymerase[3], and this interaction appears to be highly sensitive to both the distance and relative phase of the target site to the minimal promoter. The sharp phase dependence of CRISPRa may be a general feature of transcriptional regulation in *E. coli*. The native SoxS protein and other transcription factors such as CAP and LacI have restrictive positioning requirements that correspond to DNA periodicity[27–34]; we confirmed this result with an endogenous SoxS reporter (Supplementary Figure 6). In practice, this periodic behavior means that effective target sites must be located at one of the narrow peaks of activation within the optimal distance range. These stringent requirements suggest that targeting endogenous genes will be extremely challenging. There is ~1 PAM site every 10 bases in the regions upstream of endogenous promoters in *E. coli* (Supplementary Figure 7 A & B), and the

likelihood that a PAM site will be located at the appropriate phase within a 10 base window is low (Supplementary Figure 7C).

Tuning structure to expand target site range is ineffective

If rotating the CRISPRa complex out of phase along the DNA prevents SoxS from interacting with RNA polymerase, then a longer amino acid linker to SoxS might allow effective CRISPRa at more scRNA sites. To test this possibility, we extended the linker between MCP and SoxS from 5 amino acids (aa) to 10 or 20 aa, but even with these longer linkers we observed the same sharp dependence on the target site position as with the original 5 aa linker (Figure 4B). We obtained similar results using a linker with a different amino acid composition (Supplementary Figure 8A).

Another potential approach to expand the range of effective CRISPRa sites would be to change the spatial position of the MCP-SoxS protein by altering the position of the MS2 hairpin that binds MCP. We therefore tested multiple alternative scRNA designs that present the MS2 hairpin at different locations. Extending the MS2 stem by 2, 5, 10, and 20 bp resulted in progressively lower CRISPRa activity, but no change in the position of the target sites that were most effective (Supplementary Figure 8B). Similarly, no changes were observed with alternative scRNA designs with one or two MS2 hairpins presented from different locations within the scRNA structure (Supplementary Figure 8C).

Finally, we assessed whether any alternative activation domains could produce a different phase dependent behavior. Previously, these constructs all produced weaker activation than SoxS[3], perhaps because they have each distinct optimal target site positions. We tested MCP fused to TetD, αNTD, lambda cII, and RpoZ[3], and dCas9 fused to RpoZ[35]; however, none of these

constructs produced gene activation at any site that was not already effective with SoxS (Supplementary Figure 9).

Although endogenous bacterial transcription factors exhibit a sharp periodic dependence on distance[27–34], it remains surprising that no structural modifications of the CRISPRa complex produced any changes in the phase dependence. If SoxS is simply tethered to the CRISPRa complex by a flexible linker, we would have expected the peak of effective CRISPRa sites to broaden with longer linkers. The failure of this prediction suggests that our understanding of the CRISPR-Cas complex and its interactions with bacterial transcriptional machinery is fundamentally incomplete, or that the linker tethering SoxS to the CRISPRa complex is not truly flexible. Practically, it means that we still lack a way to expand the range of effective CRISPRa target sites.

A dCas9 variant expands the range of targetable sites

Because there is a limited number of genes with an appropriate NGG PAM site at precisely the optimal position upstream of the promoter (Supplementary Figure 7C), we attempted to expand the scope of targetable PAM sites for CRISPRa. We used a recently characterized dCas9 variant, dxCas9(3.7), that has improved activity at a variety of non-NGG PAM sites including NGN, GAA, GAT, and CAA[6]. We generated reporter plasmids by replacing AGG PAM sites with alternative PAM sequences and delivered a CRISPRa system with dxCas9(3.7) to target these reporters. dxCas9(3.7) maintained the ability to target the AGG PAM and showed significantly increased levels of activation at alternative PAM sites compared to dCas9 (Figure 5A). Activation levels varied with different PAM sites and correlated well with dxCas9(3.7) activity previously reported in human cells (Supplementary Figure 10A)[6]. dxCas9(3.7) showed similar distance and phase dependent target site preferences as dCas9 (Supplementary Figure 10B & C), but its expanded PAM scope makes it more likely that an arbitrary gene will have a targetable PAM site

at an effective position. Bioinformatic analysis of the sequences between transcriptional units in *E. coli* revealed that there are on average 6.4 times more dxCas9(3.7)-compatible PAM sites than NGG PAM sites (Supplementary Figure 10D). Accounting for the fact that dCas9 has some activity at non-NGG sites[6] (Figure 5A), there are still on average ~2.2-fold more dxCas9(3.7)-compatible PAM sites than dCas9-compatible PAM sites (Supplementary Figure 10D).

To demonstrate the utility of dxCas9(3.7) for CRISPRa at sites inaccessible to dCas9, we constructed a reporter plasmid that contains an AGG PAM site at the original position with maximum CRISPRa activity and an AGT PAM 5 bases downstream. Using this reporter, we observe that both dCas9 and dxCas9(3.7) are effective for CRISPRa at the optimally-positioned NGG PAM site, but neither is capable of activating the AGT PAM site, which is 5 bases out of phase from the optimal site (Figure 5B). We then inserted 5 bases into the reporter to shift the AGT PAM site into the peak activation range. With this reporter, neither dCas9 nor dxCas9(3.7) can activate the NGG PAM site, which is now out of phase. dxCas9(3.7) was now able to effectively activate the AGT PAM site, and dCas9 was ineffective at this site (Figure 5B). This result confirms that dxCas9(3.7) is able to activate optimally-positioned target sites that are inaccessible to dCas9. We expect that this behavior will be effective at many σ^{70}-family promoters (Figure 3B), and a recent report demonstrated a similar behavior of dxCas9(3.7) at σ^{54}-dependent promoters[7].

Defined rules enable endogenous gene activation

Our systematic characterization of the requirements for effective CRISPRa in *E. coli* demonstrates that candidate genes must have a targetable PAM site located at one of the sharp peaks of activity upstream of the TSS. In hindsight, the scRNA sites at endogenous genes that we initially targeted in Figure 2 did note meet this criterion. To determine if the revised rules would enable activation of endogenous *E. coli* genes, we surveyed the genome for candidate genes with

appropriately positioned, dxCas9(3.7)-compatible PAM sites (Supplementary Methods) (Supplementary Figure 7C). We selected candidates with multiple potentially effective PAM sites and further narrowed the pool based on two additional criteria: (1) genes should not be too highly expressed (Figure 3A) and (2) genes should be regulated by σ^{70}, which is the sigma factor that regulates most genes[8] (Figure 3B). Ideally, we would also exclude genes with tightly bound transcriptional regulators in the promoter region (Figure 3D), but this information is not readily available. We chose six genes that could be tested using reporter strains from the *E. coli* promoter collection[36] and targeted two PAM sites for each gene.

We first examined the *yajG* gene, which had two plausible target sites, one of which was only compatible with dxCas9(3.7). We also included an additional site predicted to be out of phase and ineffective for CRISPRa. We observed significant, ~4-6-fold gene activation for the two sites located at the predicted peak of activity at -80/-81, and no activation at the out of phase site at -87 (Figure 6A). The site at -81 is inaccessible to dCas9, and we only observed activation with dxCas9(3.7). We proceeded to test an additional five genes with partial success. We observed significant activation at *poxB* (~10-fold) and *uxuR* (~2-fold) (Figure 6B). We validated these results by performing RT-qPCR on the endogenous *yajG* and *poxB* loci. Targeting CRISPRa to these genes resulted in increases in RNA levels (Supplementary Figure 11). Targeting CRISPRa to *araE* produced a statistically significant difference in expression, but the activation measured was modest (1.13-fold). For the remaining two candidate genes, *ansB* was modestly repressed at one of the target sites and we did not observe a statistically significant difference in expression at *ppiD*. Similarly, one of the *ldhA* sites that we targeted in initial experiments (Supplementary Figure 1) was at a predicted optimal site at -91 and failed to give substantial activation. Thus, of seven endogenous genes tested with target sites that we predict should be effective (the six genes from Figure 6B and *ldhA* from Supplementary Figure 1), we were able to activate three genes with >2-fold increases in gene expression.

Although any success at endogenous gene activation is encouraging, significant challenges remain for predictable CRISPRa in bacteria. Our results suggest that even with a precise distance metric for effective target sites, some genes will not be predictably activated. There are several possible explanations: (1) tightly bound negative regulators could interfere with CRISPRa (Figure 3D), and (2) small errors in transcription start site annotation could lead to inaccurate predictions for effective sites, given that 1-2 base shifts can have dramatic effects on CRISPRa (Figure 4), and (3) intrinsic differences in the activity of the 20 base scRNA target sequence (Supplementary Figure 4).

Discussion

Bacterial CRISPRa is sensitive to a number of factors, including (i) the strength of the target promoter, (ii) the sigma factor regulating the promoter, (iii) the sequence composition immediately upstream of the minimal promoter, (iv) the composition of the scRNA target sequence, (v) the position of the scRNA target site with respect to the TSS at single base resolution. Some of these factors, such as promoter strength and scRNA target sequence composition, are also relevant in eukaryotic systems[13,15,37,38]. Other factors are plausible given our understanding of bacterial transcription. Sigma factor levels are regulated to control gene expression in response to cell state and external signals[16], so it is reasonable that we observed variable levels of activation from promoters with alternative sigma factors. Many bacterial genes are controlled by negative regulators[39], and different sequences upstream of the minimal promoter could be recruiting repressors.

The most unexpected property that we observed with bacterial CRISPRa was its sharp, periodic dependence on target site position. This behavior is quite distinct from CRISPRa in

eukaryotes, where a broad range of sites upstream of the TSS are effective[40], possibly because eukaryotic activators typically recruit transcription factors and chromatin modifying machinery rather than directly recruiting RNA polymerase. There is precedent for bacterial transcriptional activators that are sensitive to target site periodicity[27–34], but the dramatic changes in activity with only single base shifts is surprising. Moreover, it is puzzling that we were unable to predictably alter or broaden the range of sites that are effective. Our models for how activators interact with bacterial transcription machinery may be incomplete. It will likely be productive to continue screening for activity at out-of-phase target sites using additional systematic modifications to the CRISPRa complex structure, alternative CRISPR-Cas systems, and additional candidate transcriptional activation domains.

Despite the challenges described above for identifying effective CRISPRa sites in *E. coli*, our systematic characterization provides a framework for immediate practical applications and a path for future improvements. We now have a clear understanding of the criteria needed to design synthetic promoters that can be regulated by CRISPRa, which will enable the construction of complex, tunable synthetic multi-gene circuits. To extend the scope of CRISPRa to endogenous target genes, expanded PAM variants like dxCas9(3.7)[6], or orthologous dCas9 proteins with alternate PAM specificities[41,42] will open more DNA sites for targeting, increasing the likelihood of finding a targetable site at an optimal position relative to the TSS. These strategies lay the groundwork for more widespread use of bacterial CRISPRa in basic research and practical applications including functional genomics screens, metabolic engineering, and synthetic microbial communities.

Methods

Bacterial strain construction and manipulation

Plasmids were cloned using standard molecular biology protocols. Bacterial strains with sfGFP or mRFP1 reporter strains are described in Supplementary Table 1. The CRISPRa system used for each figure panel is described in Supplementary Table 2. Guide RNA target sequences are described in Supplementary Table 3. Plasmid containing the reporter genes and the CRISPR components are described in Supplementary Table 4. *S. pyogenes* dCas9 (*Sp*-dCas9) or dxCas9(3.7) were expressed from the endogenous *Sp.pCas9* promoter in a p15A vector. MCP-SoxS containing wild-type and mutant SoxS were expressed using the BBa_J23107 promoter [http://parts.igem.org] in the same plasmid with dCas9. The scRNAs were expressed using the BBa_J23119 promoter, either in the same plasmid with the dCas9 protein and the activation domain or in a separate ColE1 plasmid. The scRNA.b1 or scRNA.b2 designs, where the endogenous tracr terminator hairpin upstream of MS2 was removed[3], were used in all experiments except otherwise noted. The *zwfp-lacZ* and *fumCp-lacZ* reporter plasmids were generated in a previous study[3]. mRFP1 and sfGFP reporters were expressed from the weak BBa_J23117 minimal promoter [http://parts.igem.org] in a low-copy pSC101** vector. Variant versions of reporter genes are described in the Supplementary Methods. Plasmid libraries containing N26 sequences between the scRNA target site and BBa_J23117 minimal promoter were constructed by PCR amplification using mixed bases oligos (IDT). The dxCas9(3.7)-VPR plasmid was a gift from David Liu (Addgene #108383)[6].

Flow cytometry

Single colonies from LB plates were inoculated in 500 µL EZ-RDM (Teknova) supplemented with appropriate antibiotics and grown in 96-deep-well plates at 37 °C and shaking. Cultures were grown overnight at 37 °C and shaking and then diluted in 1:50 in DPBS and analyzed on a MACSQuant VYB flow cytometer with the MACSQuantify 2.8 software (Miltenyi Biotec). A side scatter threshold trigger (SSC-H) was applied to enrich for single cells until 10000 events were collected. The FlowJo 10.0.7 software was used to apply a narrow gate along the diagonal line

on the SSC-H vs SSC-A plot was selected to exclude the events where multiple cells were grouped together. Within the selected population, events that appeared on the edges of the FSC-A vs. SSC-A plot and the fluorescence histogram were excluded.

Plate reader experiments

Single colonies from LB plates were inoculated in 500 µL EZ-RDM (Teknova) supplemented with appropriate antibiotics and grown in 96-deep-well plates at 37 °C and shaking overnight. For experiments with the *E. coli* promoter collection[36] the activation domain was placed under the control of a tet-inducible promoter. Attempts to use constitutive CRISPRa were unsuccessful due to plasmid instability, possibly because of toxicity arising from increased expression of the target genes. Single colonies from LB plates were inoculated in 500 µL EZ-RDM supplemented with appropriate antibiotics and 400 nM anhydrotetracycline (aTc) and grown in 96-deep-well plates at 37 °C and shaking overnight. 150 µL of the overnight culture were transferred into a flat, clear-bottomed black 96-well plate and the OD_{600} and fluorescence were measured in a Biotek Synergy HTX plate reader and analyzed using the BioTek Gen5 2.07.17 software. For mRFP1 detection, the excitation wavelength was 540 nm and emission wavelength was 600 nm. For sfGFP detection, the excitation wavelength was 485 nm and emission wavelength was 528 nm.

Quantitative RT-PCR

Single colonies from LB plates were inoculated in 5 mL LB containing appropriate antibiotics and grown overnight at 37 °C and shaking. Overnight cultures were diluted 1:100 into 5 mL EZ-RDM supplemented with appropriate antibiotics and grown at 37 °C and shaking until an OD_{600} of 0.5 (using 150 µL of culture in a 96 well plate) was reached. For the experiments targeting *yajG* and *poxB*, the activation domain was placed under the control of a tet-inducible promoter and cultures in EZ-RDM were supplemented with 400 nM aTc. Cultures were pelleted and total RNA was

extracted using the Aurum Total RNA Mini Kit (Bio-rad). Reverse transcription reactions were performed from 1 µg RNA in 20 µL reactions using iScript reverse transcriptase (Bio-Rad). qPCR reactions were prepared in triplicate in a final volume of 10 µL using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad), 0.5-5 ng of cDNA and 400 nM primers. The reaction was performed in a CFX Connect (Bio-Rad) with a 58 °C annealing temperature and 30 s extension time. A list of the qPCR primer sequences is provided in Supplementary Table 5. Expression levels for each gene were calculated in the Bio-Rad CFX Maestro 4.0.23225.0418 software by normalizing to the 16S rRNA gene and relative to a negative control carrying an off target-scRNA using the ΔΔCT method[43].

Statistics and reproducibility

Statistical significance was calculated using two-tailed unpaired Welch's t-tests. To ensure reproducibility, experiments were performed using n = 3 biologically independent samples, unless otherwise noted.

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this article is available as a Supplementary Information file. The datasets generated and analyzed during the current study are available from the corresponding author upon request. The source data underlying Figures 1B, 2A-B, 3A-D, 4A-B, 5A-B, 6A-B and Supplementary Figures 1, 2, 3A-B, 4 , 5A-F, 6, 7A-C, 8A-C, 9, 10A-D, 11 are provided as a Source Data file.

Code availability

Custom Python code to generate the DNA sequences between transcriptional units in *E. coli* and analyze the density of PAM sites in these sequences (detailed in the Supplementary Methods) is available on GitHub [https://github.com/carothersresearch/Fontana-Dong_2020_NatComm].

References

1. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat Methods* **11**, 508–520 (2014).
2. Nielsen, J. & Keasling, J. D. Engineering cellular metabolism. *Cell* **164**, 1185–1197 (2016).
3. Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nat Commun* **9**, 2489 (2018).
4. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
5. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in genome gditing and beyond. *Annual Review of Biochemistry* **85**, 227–264 (2016).
6. Johnny H. Hu, D. R. L., Shannon M. Miller, Maarten H. Geurts, Weixin Tang, Liwei Chen, Ning Sun, Christina M. Zeina, Xue Gao, Holly A. Rees, Zhi Lin. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
7. Yang Liu, B. W., Xinyi Wan. Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria. *Nature Communications* **10**, 3693 (2019).
8. Keseler, I. M. *et al.* The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res* **45**, D543–D550 (2017).
9. Rodriguez, A. *et al.* Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. *Microb Cell Fact* **13**, 126 (2014).
10. C R Byrne, N. M. K., R. S. Monroe, K. A. Ward. DNA sequences of the *cysK* regions of *Salmonella typhimurium* and *Escherichia coli* and linkage of the *cysK* regions to *ptsH*. *Journal of Bacteriology* **170**, 3150–3157 (1988).
11. Gene Ruijun Jiang, D. P. C., Sonia Nikolova. Regulation of the *ldhA* gene, encoding the fermentative lactate dehydrogenase of *Escherichia coli*. *Microbiology* **147**, 2437–2446 (2001).

12. Olin K. Silander, M. A., Nela Nikolic, Alon Zaslaver, Anat Bren, Ilya Kikoin, Uri Alon. A genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*. *PLoS Genetics* **8**, 5 (2012).

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

14. Alejandro Chavez, G. C., Marcelle Tuttle, Benjamin W. Pruitt, Ben Ewen-Campen, Raj Chari, Dmitry Ter-Ovanesyan, Sabina J. Haque, Ryan J. Cecchi, Emma J. K. Kowal, Joanna Buchthal, Benjamin E. Housden, Norbert Perrimon, James J. Collins. Comparison of Cas9 activators in multiple species. *Nature Methods* **13**, 563–567 (2016).

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

16. Gruber, T. M. & Gross, C. A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57**, 441–466 (2003).

17. Ishita M. Shah, R. E. W. Novel protein–protein interaction between *Escherichia coli* SoxS and the DNA binding determinant of the RNA polymerase α subunit: SoxS functions as a co-sigma factor and redeploys RNA polymerase from UP-element-containing promoters to SoxS-dependent promoters during oxidative stress. *Journal of Molecular Biology* **343**, 513–532 (2004).

18. Amy Strohmeier Gort, J. A. I., Daniel M. Ferber. The regulation and role of the periplasmic copper, zinc superoxide dismutase of *Escherichia coli*. *Molecular Microbiology* **32**, 179–191 (1999).

19. Gen Nonaka, V. A. R., Matthew Blankschien, Christophe Herman, Carol A. Gross. Regulon and promoter analysis of the *E. coli* heat-shock factor, σ^{32}, reveals a multifaceted cellular response to heat stress. *Genes & Development* **20**, 1776–1789 (2006).

20. Virgil A Rhodius, C. A. G., Won Chul Suh, Gen Nonaka, Joyce West. Conserved and variable functions of the σ^{E} stress response in related genomes. *PLoS Biology* **4**, e2 (2006).

21. Zhe-Xian Tian, Y. W., Quan-Sheng Li, Martin Buck, Annie Kolb. The CRP–cAMP complex and downregulation of the *glnAp2* promoter provides a novel regulatory linkage between carbon metabolism and nitrogen assimilation in *Escherichia coli*. *Molecular Microbiology* **41**, 911–924 (2001).

22. Avihu H. Yona, J. G., Eric J. Alm. Random sequences rapidly evolve into de novo promoters. *Nature Communications* **9**, 1530 (2018).

23. Lee, T. *et al.* BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Eng* **5**, 12 (2011).

24. Alfredo Mendoza-Vargas, E. M., Leticia Olvera, Maricela Olvera, Ricardo Grande, Leticia Vega-Alvarado, Blanca Taboada, Verónica Jimenez-Jacinto, Heladia Salgado, Katy Juárez, Bruno Contreras-Moreira, Araceli M. Huerta, Julio Collado-Vides. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS ONE* **4**, e7526 (2009).

25. Georgina Lloyd, S. B., Paolo Landini. Activation and repression of transcription initiation in bacteria. *Essays In Biochemistry* **37**, 17–31 (2001).

26. Schneider, T. D. 70% efficiency of bistate molecular machines explained by information theory, high dimensional geometry and evolutionary convergence. *Nucleic Acids Res.* **38**, 5995–6006 (2010).

27. Timothy I. Wood, R. E. W., Kevin L. Griffith, William P. Fawcett, Kam-Wing Jair, Thomas D. Schneider. Interdependence of the position and orientation of SoxS binding sites in the transcriptional activation of the class I subset of *Escherichia coli* superoxide-inducible promoters. *Molecular Microbiology* **34**, 414–430 (1999).

28. Yi Zhou, Y.-P. W., Annie Kolb, Stephen J. W. Busby. Spacing requirements for Class I transcription activation in bacteria are set by promoter elements. *Nucleic Acids Research* **42**, 9209–9216 (2014).

29. Johannes Müller, B. M.-H., Andrew Barker, Stefan Oehler. Dimeric *lac* repressors exhibit phase-dependent co-operativity. *Journal of Molecular Biology* **284**, 851–857 (1998).

30. Straney, D. C., Straney, S. B. & Crothers, D. M. Synergy between *Escherichia coli* CAP protein and RNA polymerase in the *lac* promoter open complex. *J. Mol. Biol.* **206**, 41–57 (1989).

31. Gaston, K., Bell, A., Kolb, A., Buc, H. & Busby, S. Stringent spacing requirements for transcription activation by CRP. *Cell* **62**, 733–743 (1990).

32. Ushida, C. & Aiba, H. Helical phase dependent action of CRP: effect of the distance between the CRP site and the −35 region on promoter activity. *Nucleic Acids Res* **18**, 6325–6330 (1990).

33. Müller, J., Oehler, S. & Müller-Hill, B. Repression of *lac* Promoter as a function of distance, phase and quality of an auxiliary *lac* operator. *Journal of Molecular Biology* **257**, 21–29 (1996).

34. Martin, R. G., Gillette, W. K., Rhee, S. & Rosner, J. L. Structural requirements for marbox function in transcriptional activation of *mar*/*sox*/*rob* regulon promoters in *Escherichia coli*: sequence, orientation and spatial relationship to the core promoter. *Mol. Microbiol.* **34**, 431–441 (1999).

35. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **41**, 7429–7437 (2013).

36. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods* **3**, 623–628 (2006).

37. Kiani, S. *et al.* CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat Methods* **11**, 723–726 (2014).

38. Gander, M. W., Vrana, J. D., Voje, W. E., Carothers, J. M. & Klavins, E. Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nat Commun* **8**, 15459 (2017).

39. Rojo, F. Repression of transcription initiation in bacteria. *Journal of bacteriology* **181**, 2987–91 (1999).

40. Gilbert, L. A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).

41. Ryan T. Leenay, C. L. B. Deciphering, communicating, and engineering the CRISPR PAM. *Journal of Molecular Biology* **429**, 177–191 (2017).

42. Sergey Shmakov, E. V. K., Aaron Smargon, David Scott, David Cox, Neena Pyzocha, Winston Yan, Omar O. Abudayyeh, Jonathan S. Gootenberg, Kira S. Makarova, Yuri I. Wolf, Konstantin Severinov, Feng Zhang. Diversity and evolution of class 2 CRISPR-Cas systems. *Nature Reviews Microbiology* **15**, 169–182 (2017).

43. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).

44. Griffith, K. L. & Wolf, R. E. A comprehensive alanine scanning mutagenesis of the *Escherichia coli* transcriptional activator SoxS: Identifying amino acids important for DNA binding and transcription activation. *Journal of Molecular Biology* **322**, 237–257 (2002).

Acknowledgements

The dxCas9(3.7)-VPR plasmid was a gift from David Liu (Addgene #108383). The authors thank Mary Lidstrom, Joanne Wong, Semira Beraki and members of the Zalatan and Carothers groups for technical assistance, advice, and helpful discussions. This work was supported by NSF Award 1817623 (J.M.C, J.G.Z.) and NSF Award 1844152 (J.M.C.).

Author Contributions

J.F., C.D., C.K., B.I.T., J.M.C., and J.G.Z. designed experiments and analyzed data. J.F., C.D., C.K., V.P.C., and B.I.T. performed experiments. J.F., C.D., J.M.C. and J.G.Z. wrote the manuscript.

Competing Interests

The authors declare no competing interests

Figure Legends

Figure 1. A SoxS double mutant maintains CRISPRa activity and does not activate endogenous SoxS targets.

A) Reporter system for measuring the CRISPRa activity and endogenous SoxS-dependent gene expression of wild-type or mutant SoxS constructs. CRISPRa activity was determined in a strain harboring a genomically-integrated sfGFP reporter (CD06, Supplementary Table 1). The endogenous SoxS-dependent gene expression was determined by monitoring *lacZ* expression from reporter plasmids where *lacZ* was driven by SoxS-regulated promoters *zwfp* and *fumCp*[44]. GFP fluorescence was measured by flow cytometry and *lacZ* activity was measured using a β-galactosidase assay. **B)** SoxS(R93A/S101A) maintains CRISPRa activity and does not activate expression from the endogenous expression from the *zwfp* and *fumCp* reporters. Fluorescence and *lacZ* activity values were baseline-subtracted using a strain that does not express a scRNA. Both GFP levels and *lacZ* activities were normalized to the values observed in the strain with wild-type SoxS. Values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data for panel B are provided as a Source Data file.

Figure 2. A simple distance metric does not predict CRISPRa activity.

A) CRISPRa on the *aroK-aroB* operon. Two scRNA target sites within the 40 base window where CRISPRa is effective (-100 to -60) in heterologous reporter genes (A3-A4) and two sites further upstream (A1-A2) were chosen for the *aroKp1* promoter. **B)** CRISPRa on the *cysK* gene. Three scRNA target sites within the 40 base window where CRISPRa is effective in heterologous reporter genes (C1-C3) and two sites further downstream (C4-C5) were chosen for the *cysKp2* promoter. The C4 and C5 sites resulted in repression; targeting these sites close to the core promoter may interfere with RNA polymerase binding. Gene expression was measured using RT-qPCR. Fold activation represents expression levels relative to a strain expressing an off-target scRNA (hAAVS1). In panels A and B, bars represent the average +/- standard deviation calculated from n = 4 (C1, C2), n = 3 (A1-A4, C4,C5) or n = 2 (C3) biologically independent samples. Some individual replicates in samples A1-A4 and C1-C2 appear to show activation, but a two-tailed unpaired Welch's t-test indicates that the average differences relative to the off-target control are not statistically significant (p-value > 0.05). Targeting CRISPRa to C4-C5 resulted in repression, likely due the short distance between the sites and the promoter. Stars indicate a statistically significant difference from the off-target control using a two-tailed unpaired Welch's t-test (**: p-value < 0.01). Exact p-values: A1: 0.18, A2: 0.27, A3: 0.36, A4: 0.17, C1: 0.29, C2: 0.66, C3: 0.098, C4: 0.0022, C5: 0.00043. Source data for panels A and B are provided as a Source Data file.

Figure 3. CRISPRa is sensitive to promoter identity and local sequence.

A) CRISPRa is sensitive to promoter strength. Promoters contain a scRNA target site at -81 from the TSS of the indicated J231NN minimal promoter, on the non-template strand[3]. The panel on the left shows the Fluorescence/OD_{600} of strains expressing an on-target or off-target scRNA. The panel on the right shows the fold activation measured at each promoter relative to their baseline expression with an off-target scRNA (J206). **B)** CRISPRa can activate promoters regulated by σ^{38} (RpoS), σ^{32} (RpoH), and σ^{24} (RpoE) sigma factors. The minimal promoter from the reporter plasmid was replaced with *sodCp*, *glnAp2*, *rdgBp*, or *yieEp*. The -35 and -10 regions are highlighted in bold. The plot on the left shows the Fluorescence/OD_{600} when CRISPRa targeted each promoter at the J109 target site (-80 from the TSS on the template strand) or with an off-target scRNA (hAAVS1, labeled (–)). The plot on the right shows the fold activation measured at each promoter relative to an off-target scRNA (J206). **C)** CRISPRa activity differs significantly among promoters with varying sequence composition between the scRNA target and the -35 region. Green bars represent the Fluorescence/OD_{600} of overnight cultures from individual colonies. The blue bar represents the Fluorescence/OD_{600} of a strain expressing the J3-J23117-sfGFP reporter, activated by CRISPRa with the J306 scRNA. The grey bar represents a negative control expressing the J3-J23117-sfGFP reporter plasmid with CRISPRa targeting an off-target site (J206). **D)** CRISPRa was inhibited binding of the TetR transcriptional repressor binding to a tet operator (tetO) site placed upstream of the -35 region. Cultures where CRISPRa was targeted to the J306 site or to an off-target site (J206) were grown overnight in media +/- 1 μM aTc. In panels A, B and D, values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. In panel C, bars represent the value of n = 1 biologically independent samples. Source data for panels A, B, C and D are provided as a Source Data file.

Figure 4. CRISPRa is sensitive to the precise position of the scRNA target.

A) CRISPRa displays periodic positioning dependence with peak activities every 10-11 bases between -60 to -100 from the TSS. Reporter genes were constructed by inserting 0-12 bases upstream of the -35 region of the J1-J23117-mRFP1 reporter. Five scRNA sites (J102, J104, J106, J108, J110) with positions -61, -71, -81, -91, -101 from the TSS on the non-template strand of the original promoter were targeted. In this way, the complete -61 to -113 region can be covered at single base resolution. The color coding indicates data for the same target site shifted across a 12 base window. The panel on the right shows the baseline expression of reporters with shifted bases when an off-target scRNA was used (J206). The grey area represents the range of the baselines among the reporter series. For comparison, previous CRISPRa data for the J102, J104, J106, J108, J110 target positions at 10 base resolution are shown on the schematic above the plot[3]. **B)** Extending the linker length between MCP and SoxS does not change the position dependence of CRISPRa. The J1-J23117-mRFP1 reporter plasmid series with base shifts were delivered together with CRISPRa components for targeting J106. The MCP-SoxS(R93A) effector contained 5aa, 10aa, and 20aa linkers. For comparison, previous CRISPRa data with 5aa, 10aa, or 20aa linker between MCP and SoxS targeting at the -81 and -91 positions are shown on the schematic above the plot[3]. Values in panels A and B represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data for panels A and B are provided as a Source Data file.

Figure 5. dxCas9(3.7) expands the range of targetable scRNA target sites by recognizing alternative PAMs.

A) CRISPRa with dxCas9(3.7) displayed activity on non-NGG PAM sites with AGA, AGC, AGT, CGA, CGC, CGT, GGA, GGC, GGT, TGA, TGC, TGT, GAA, GAT, CAA sequences. CRISPRa activity with dxCas9(3.7) on non-NGG PAM sites was generally lower (6-fold to 89-fold activation relative to a control without a scRNA) compared to the AGG PAM site (188-fold activation). *Sp*-dCas9 also displayed moderate CRISPRa activity at non-NGG PAM sites with AGA, CGA, GGA, GGC, GGT, TGA sequences, consistent with published reports[6]. Reporter plasmids were constructed by replacing the AGG PAM site for the J306 target in the J3-J23117-mRFP1 reporter with alternative PAM sequences that have been previously reported to be recognized by dxCas9(3.7) in human cells[6]. The (-) sign indicates a control expressing the original reporter with the AGG PAM and the CRISPRa components with *Sp*-dCas9, the activation domain and no scRNA. **B)** dxCas9(3.7) can activate promoters that cannot be activated by *Sp*-dCas9. When the scRNA target at the optimal position (-81 to the TSS) has an AGG PAM site, both *Sp*-dCas9 and dxCas9(3.7) increased gene expression by 50-fold. When the scRNA target at the optimal position has an AGT PAM site, only dxCas9(3.7) displayed a 7-fold increase in gene expression while *Sp*-dCas9 was inactive. The reporter gene has a target with an AGG PAM (M1) and a target with an AGT PAM (M2) upstream of a BBa_J23117 minimal promoter. In reporter gene A, the AGG target was located -81 to the TSS on the non-template strand and the AGT target was located -76 to the TSS on the non-template strand. In reporter gene B, 5 bases were inserted upstream of the -35 region, shifting the locations of the AGG target and AGT target to -86 and -81, respectively. The (-) sign indicates a negative control strain that contains the reporter plasmid and a plasmid expressing Sp-dCas9, the activation domain and an off-target scRNA (J206). Bars in panels A and B represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data for panels A and B are provided as a Source Data file.

Figure 6. Predictive rules enable endogenous activation.

A) CRISPRa using dCas9 and dxCas9(3.7) was targeted to a *yajG* reporter plasmid from the *E. coli* promoter collection[36]. Three scRNA target sites were selected; two sites were located at the positions where CRISPRa was most effective (Y1-Y2), and one was located out of phase (Y3). A negative control (OT) expressing an off-target scRNA (J306) was included. **B)** CRISPRa was targeted to *yajG* and five additional promoters from the *E. coli* promoter collection (Supplementary Methods). Two scRNA sites located at the positions where CRISPRa was most effective were targeted for each gene using dxCas9(3.7). Samples are arranged by baseline expression of the target genes, in ascending order left to right. Fold activation indicates the median fluorescence of strains relative to an off-target control (J306). Values in panels A and B represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Stars indicate a statistically significant difference from the off-target control using a two-tailed unpaired Welch's t-test (*: *p*-value < 0.05, **: *p*-value < 0.01). Exact *p*-values: E1: 0.036, E2: 0.024, B1: 0.031, B2: 0.141, P1: 0.033, P2: 0.021, D1: 0.088, D2: 0.585, U1: 0.0008, U2: 0.001, Y1: 0.013, Y2: 0.003. Source data for panels A and B are provided as a Source Data file.

Figure 1: A SoxS double mutant does not activate endogenous SoxS targets

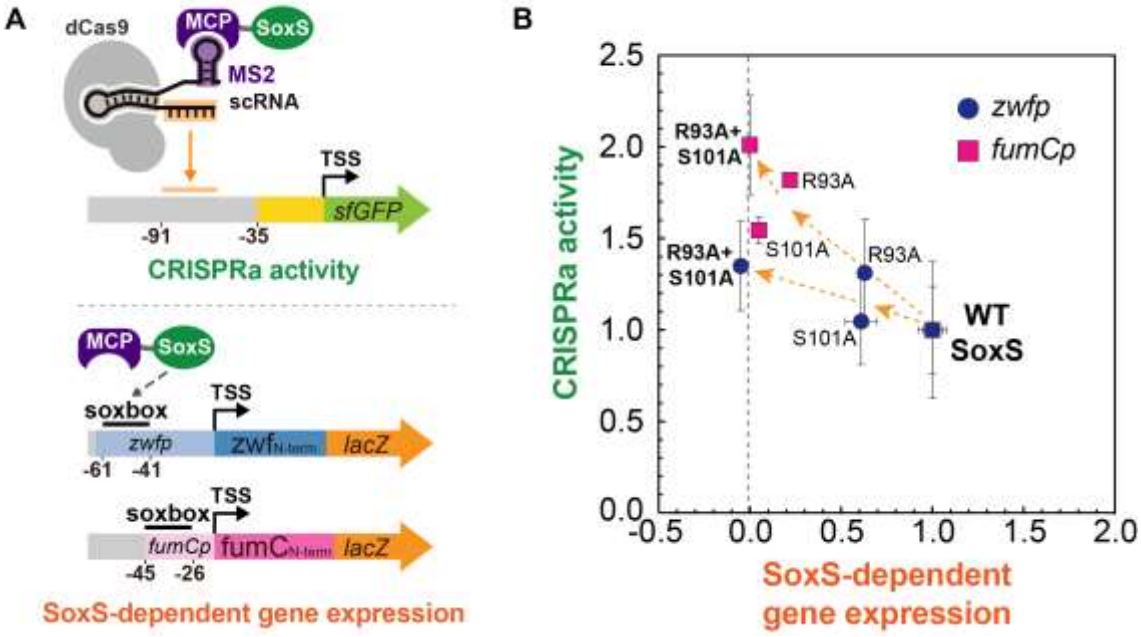


Figure 2: A simple distance metric does not predict CRISPRa activity

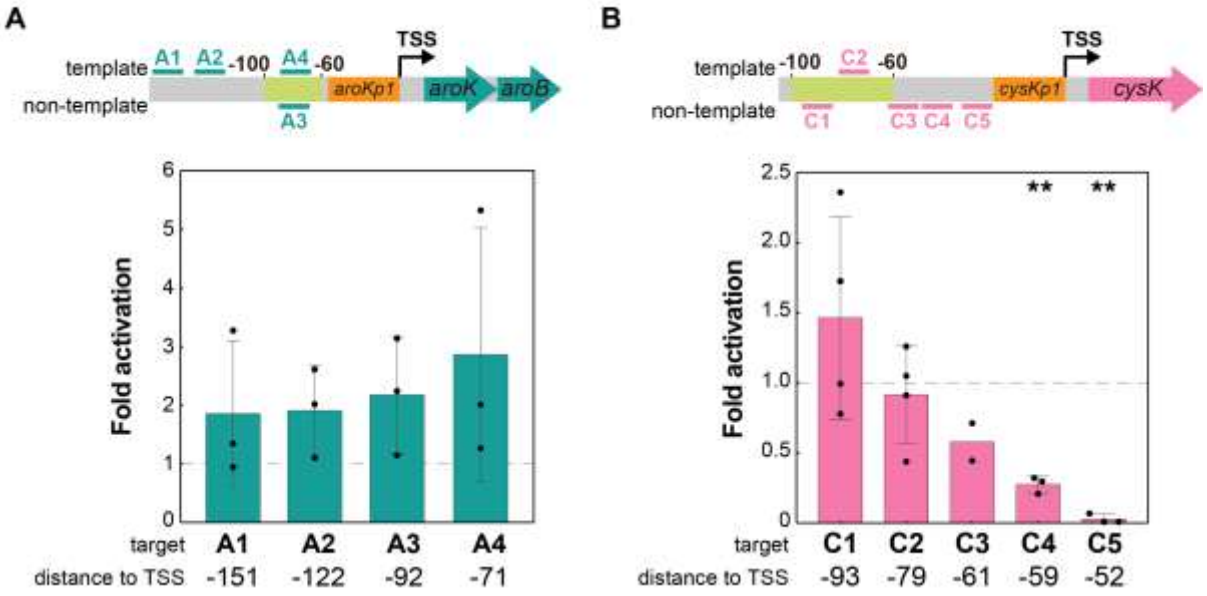


Figure 3: CRISPRa is sensitive to promoter identity and local sequence

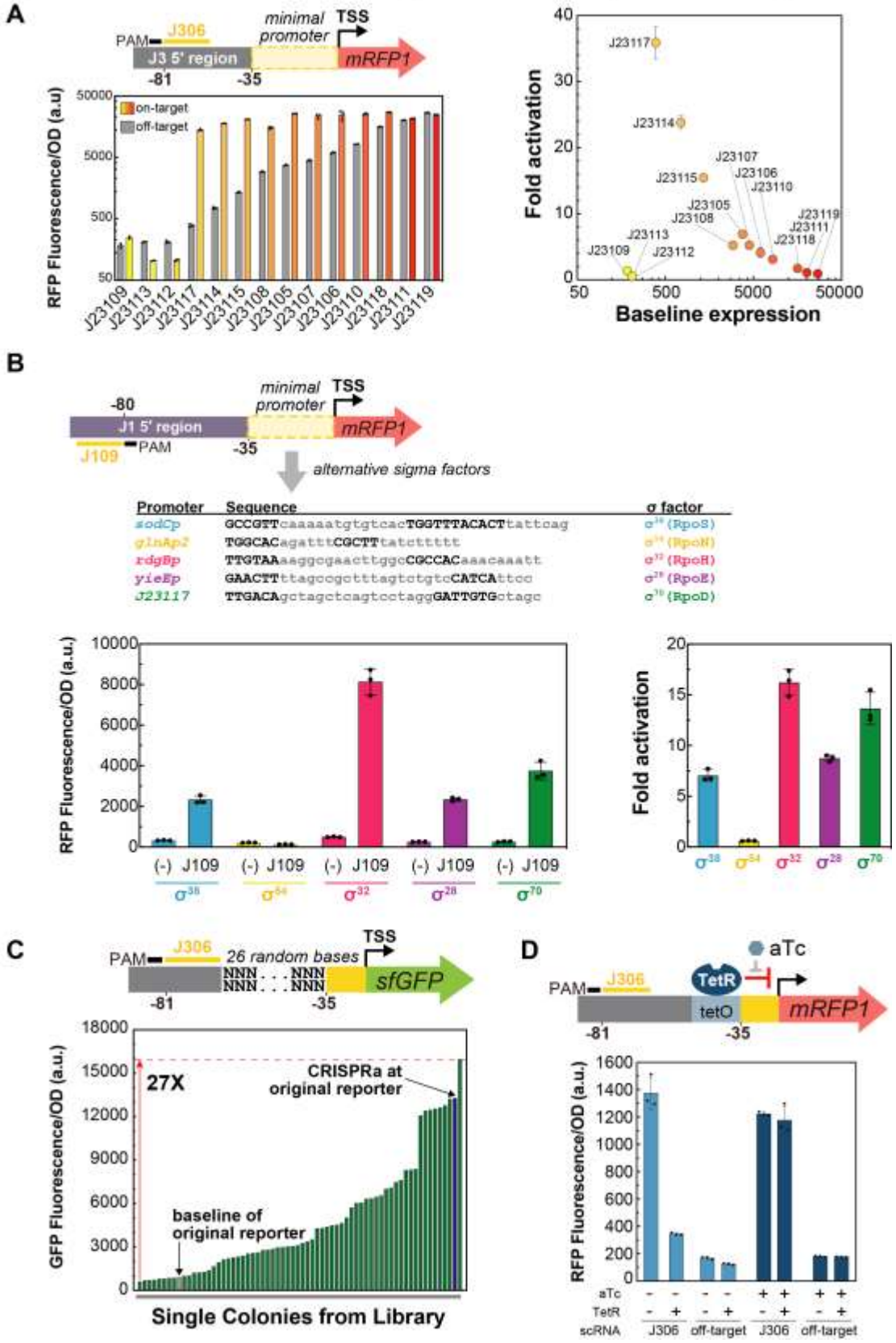


Figure 4: CRISPRa is sensitive to the precise position of the scRNA target

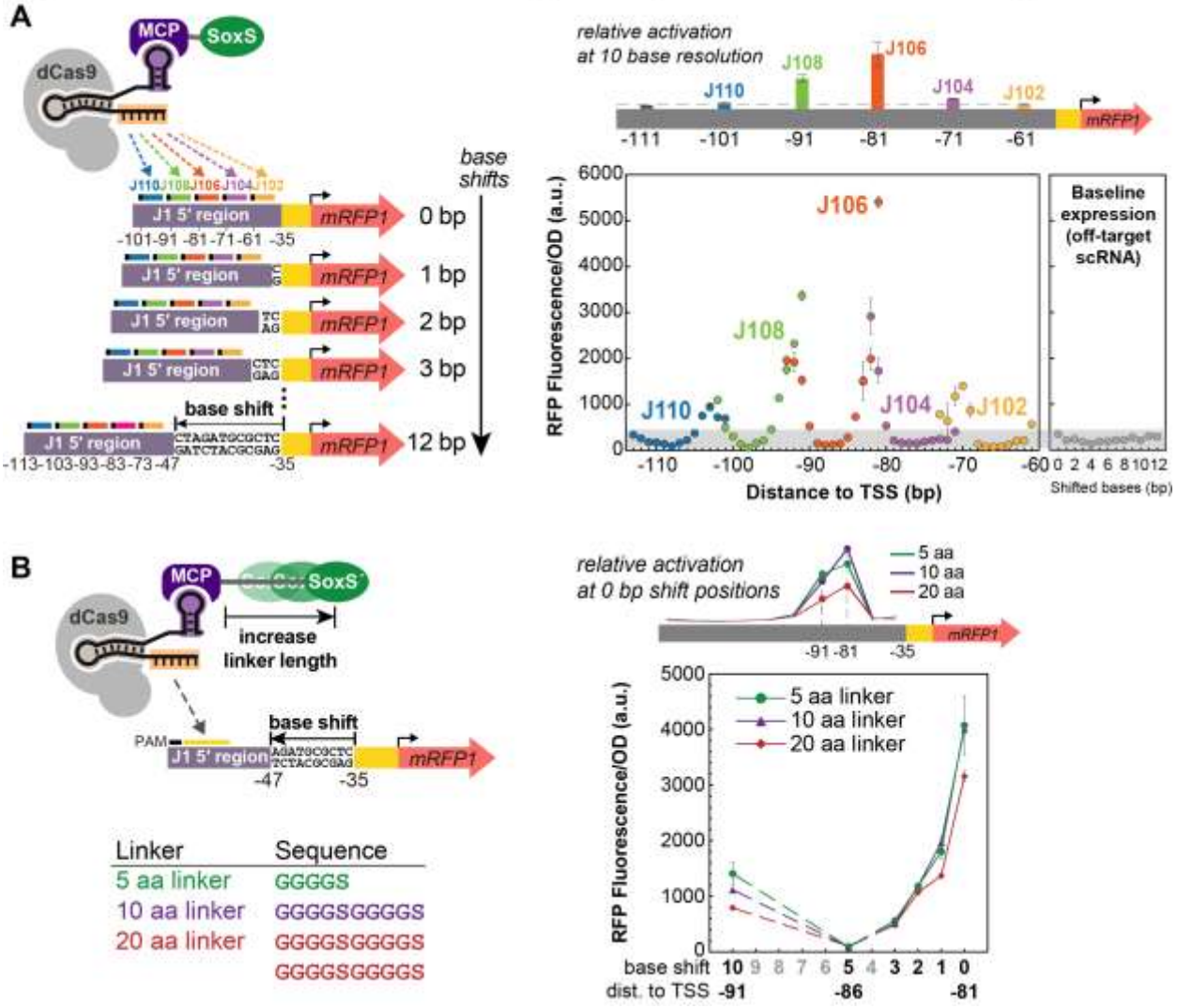


Figure 5: dxCas9(3.7) expands the range of targetable sites

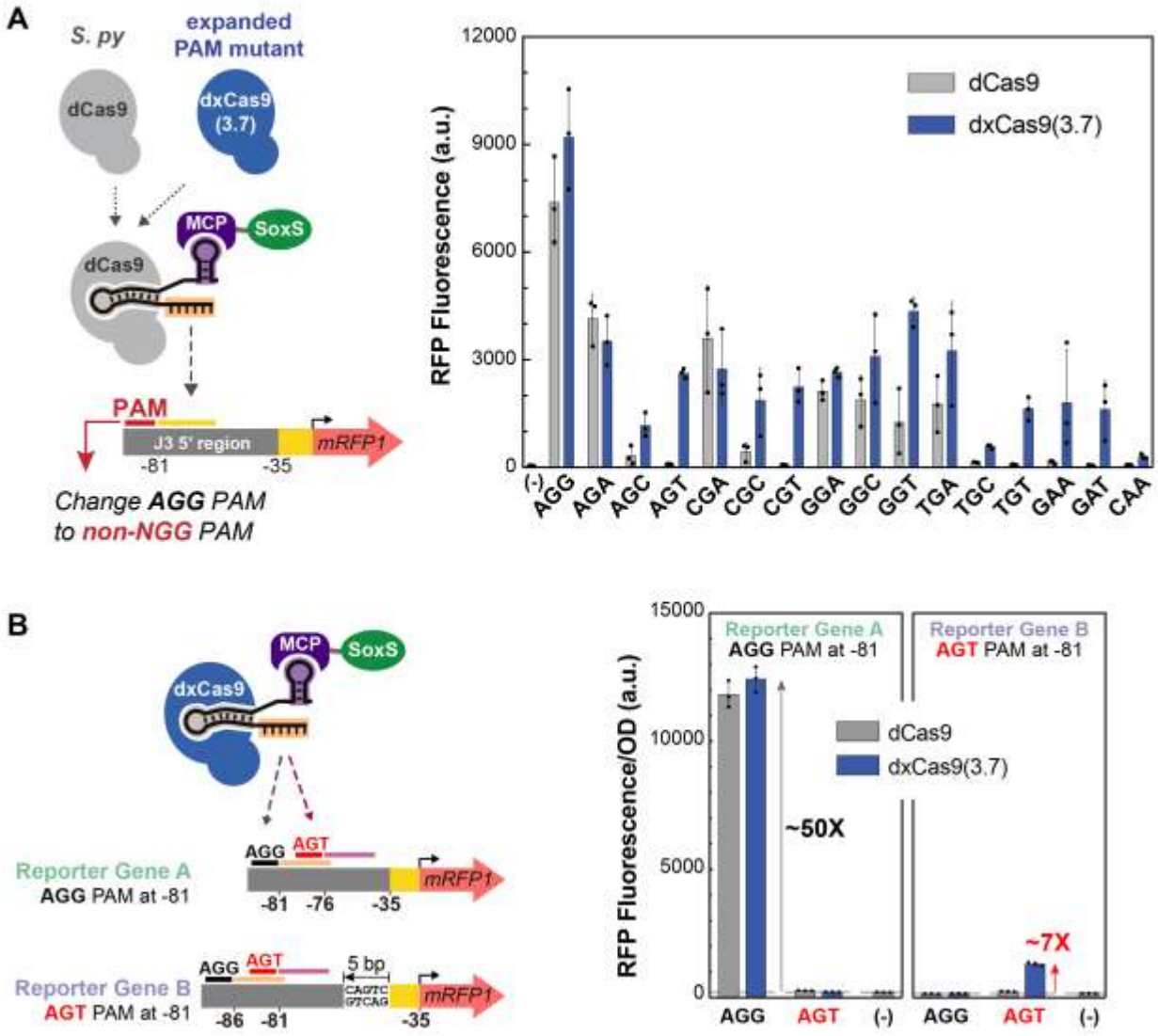
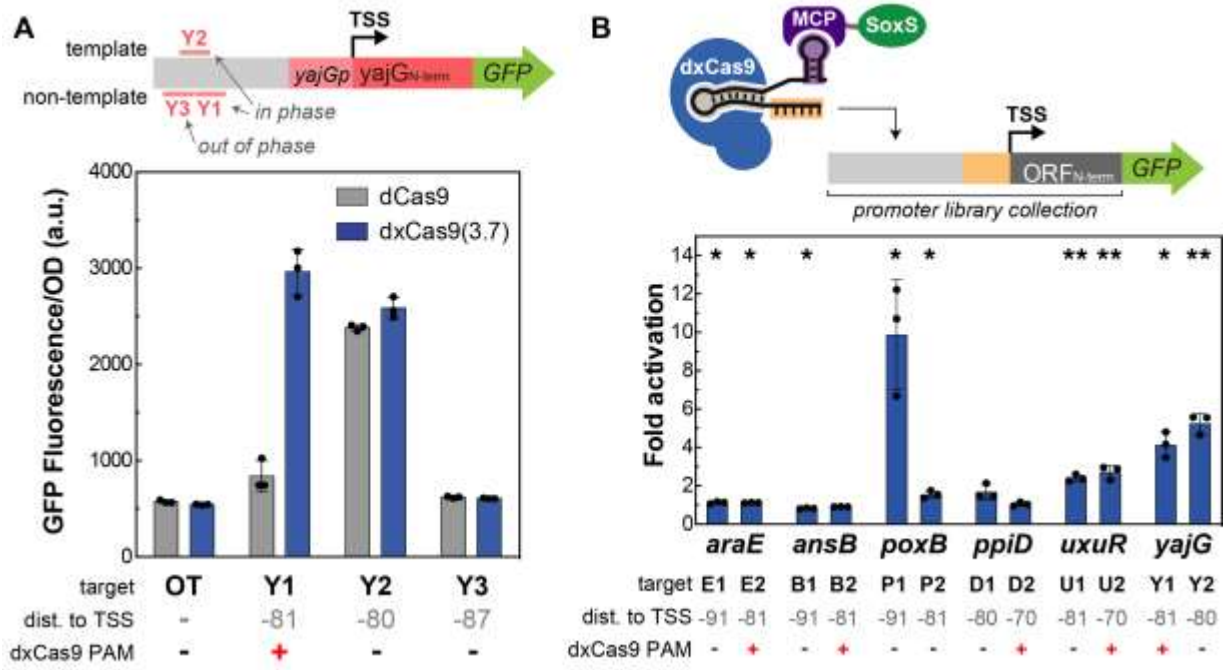


Figure 6: Predictive rules enable endogenous gene activation



SUPPLEMENTARY INFORMATION

Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements

Jason Fontana*, Chen Dong*, Cholpisit Kiattisewee, Venkata P. Chavali, Benjamin I. Tickman, James M. Carothers+, Jesse G. Zalatan+

* these authors contributed equally
+ To whom correspondence should be addressed: zalatan@uw.edu or jcaroth@uw.edu

[1]Molecular Engineering & Sciences Institute
[2]Department of Chemistry
[3]Department of Chemical Engineering
[4]Center for Synthetic Biology
University of Washington, Seattle, WA 98195, USA

This PDF file includes:

Supplementary Figures 1-11

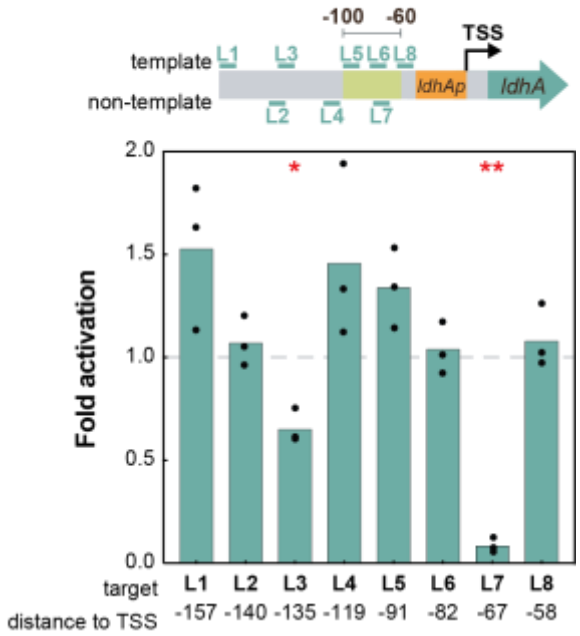
Supplementary Tables 1-6

Supplementary Methods, including reporter gene sequences

Supplementary References

Supplementary figures

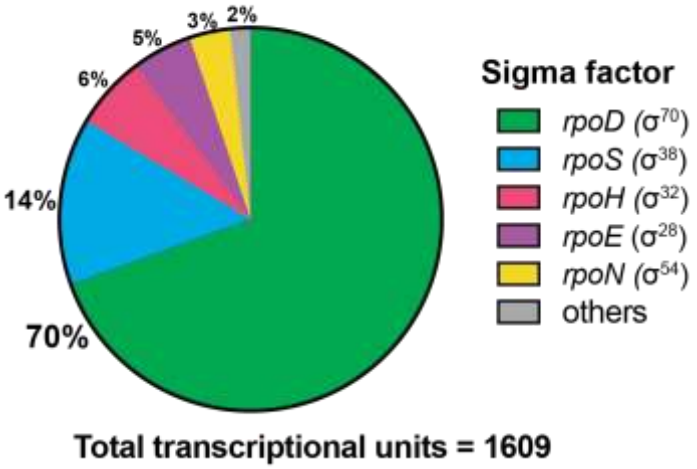
Figure S1



Supplementary Figure 1: CRISPRa at the endogenous gene target *ldhA* does not follow predicted trends.

Eight scRNA target sites (L1-L8) upstream of the *ldhA* promoter were selected. Three of the target sites (L5-L7) were within the 40 bp window where CRISPRa is effective (-100 to -60). While L1, L4, and L5 resulted in weak increases in gene expression, there was no apparent relationship between the position of the sites and *ldhA* expression levels. Gene expression was measured using RT-qPCR. Fold activation represents expression levels relative to an off-target control (hAAVS1). Values represent the average calculated from n = 3 technical replicates. Stars indicate a statistically significant difference from the off-target control using a two-tailed unpaired Welch's t-test (*: *p-value* < 0.05, **: *p-value* < 0.01). Exact *p-values*: L1: 0.12, L2: 0.43, L3: 0.02, L4: 0.20, L5: 0.09, L6: 0.67, L7: 0.0005, L8: 0.45. Source data are provided as a Source Data file.

Figure S2

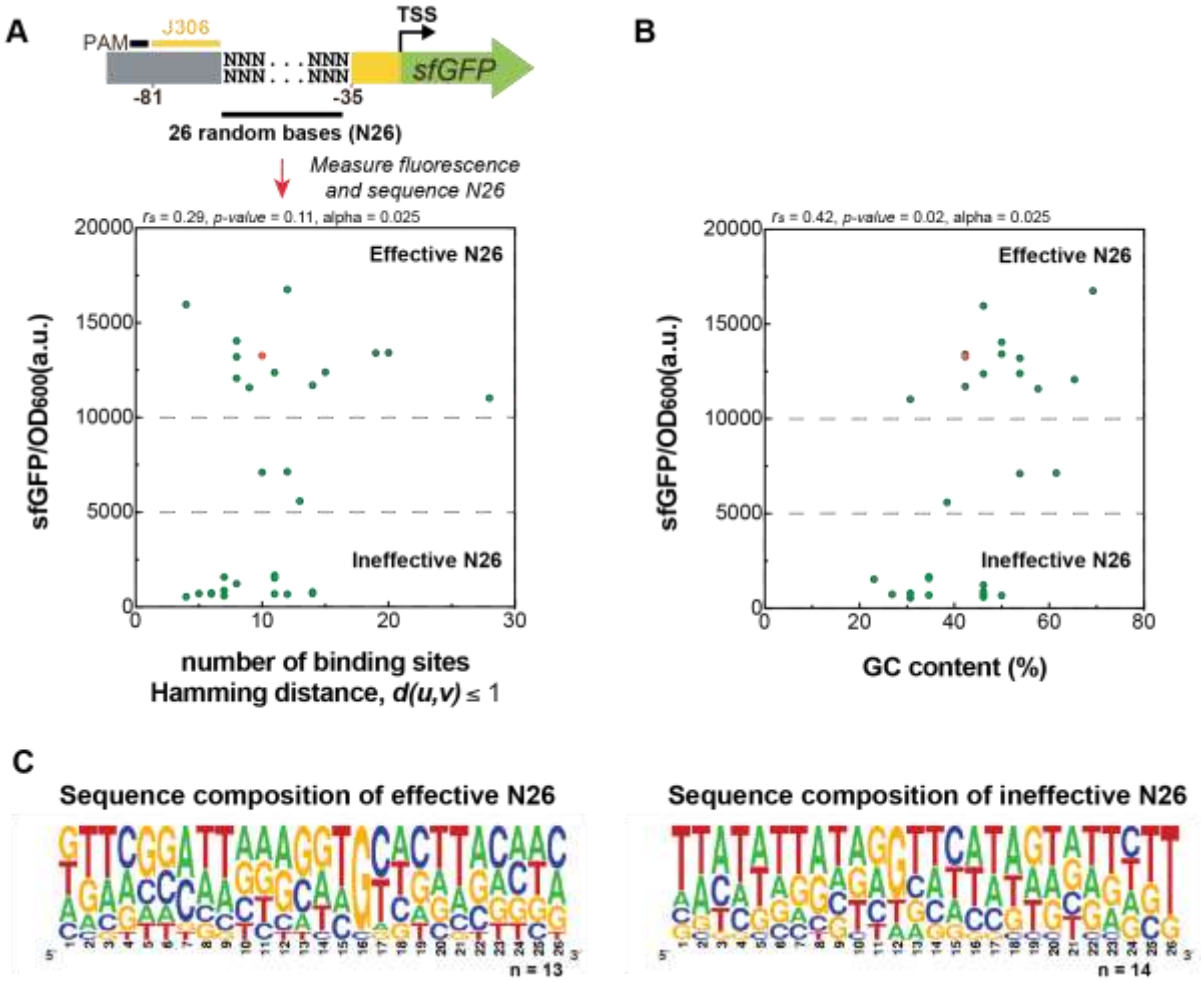


Total transcriptional units = 1609

Supplementary Figure 2: Distribution of transcriptional units regulated by sigma factors.

The number of *E. coli* transcriptional units regulated by sigma factors was obtained from Ecocyc[1] on the "Regulon" tab of the page relative to each sigma factor. The total number of transcriptional units represents the sum of the transcriptional units regulated by each sigma factor. Source data are provided as a Source Data file.

Figure S3

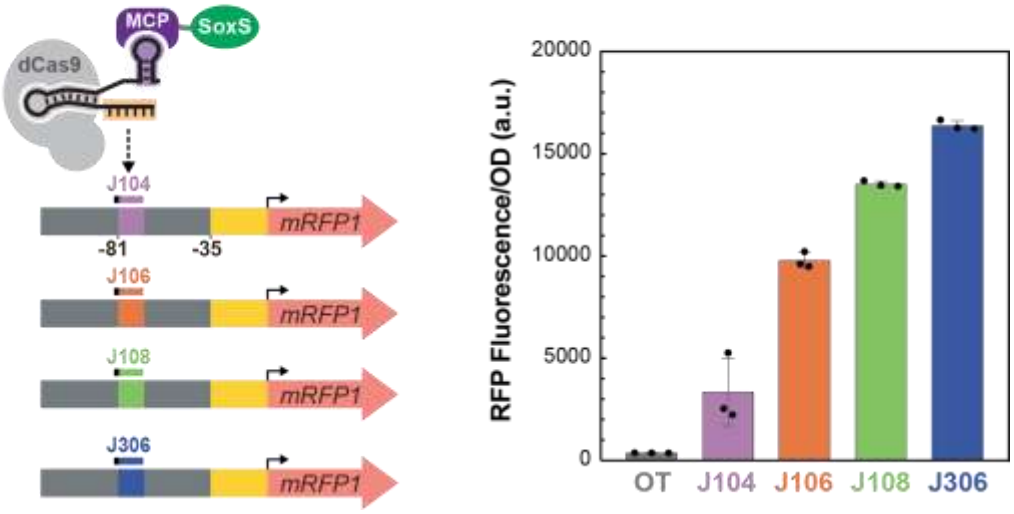


Supplementary Figure 3: Intervening sequences between the scRNA target site and minimal promoter that interfere with CRISPRa tend to be more AT-rich.

A) Plot of gene activation vs. transcription factor binding sites present in the intervening sequence between the scRNA target site and the minimal promoter. CRISPRa was targeted to a reporter library with 26 randomized bases between the scRNA target site and the -35 region on the J3-J23117-sfGFP reporter (N26), and we observed a 27-fold variation in gene activation (Figure 3C). We sequenced 29 variants and identified a number of motifs within one base of known consensus transcription factor binding (Supplementary Table 6). In parallel, we measured gene activation (sfGFP/OD_{600}) for each strain. The plot shows the sfGFP/OD_{600} of each strain versus the number of sequences within a Hamming distance $d(u,v) \leq 1$ from the consensus sequences of transcription factor binding sites, obtained from RegulonDB[2]. r_s indicates the Spearman rank order correlation coefficient between sfGFP/OD_{600} and number of binding sites, and its associated two-tailed *p-value* is relative to the null hypothesis of no correlation between sfGFP/OD_{600} and number of binding sites, with a Bonferroni-corrected $\alpha = 0.025$. Green dots indicate the sfGFP/OD_{600} values and the

number of binding sites calculated from individual colonies, and the orange dot indicates a strain where CRISPRa is targeting the original J3-J23117-sfGFP promoter. **B)** Plot of gene activation vs. GC content in the intervening sequence between the scRNA target site and the minimal promoter. r_s indicates the Spearman rank order correlation coefficient between sfGFP/OD_{600} and GC content, and its associated two-tailed *p-value* is relative to the null hypothesis of no correlation between sfGFP/OD_{600} and GC content, with a Bonferroni-corrected $\alpha = 0.025$. In panels A and B, green dots indicate the sfGFP/OD_{600} values and GC content calculated from n = 1 biologically independent samples, and the orange dot indicates a strain where CRISPRa targets the original J3-J23117-sfGFP promoter. **C)** Logo plots (https://weblogo.berkeley.edu/) showing the base composition of N26 sequences effective for CRISPRa (sfGFP/OD_{600} > 10000 a.u., 13 sequences) and N26 sequences ineffective for CRISPRa (sfGFP/OD_{600} < 5000 a.u., 14 sequences). Source data for panels A and B are provided as a Source Data file.

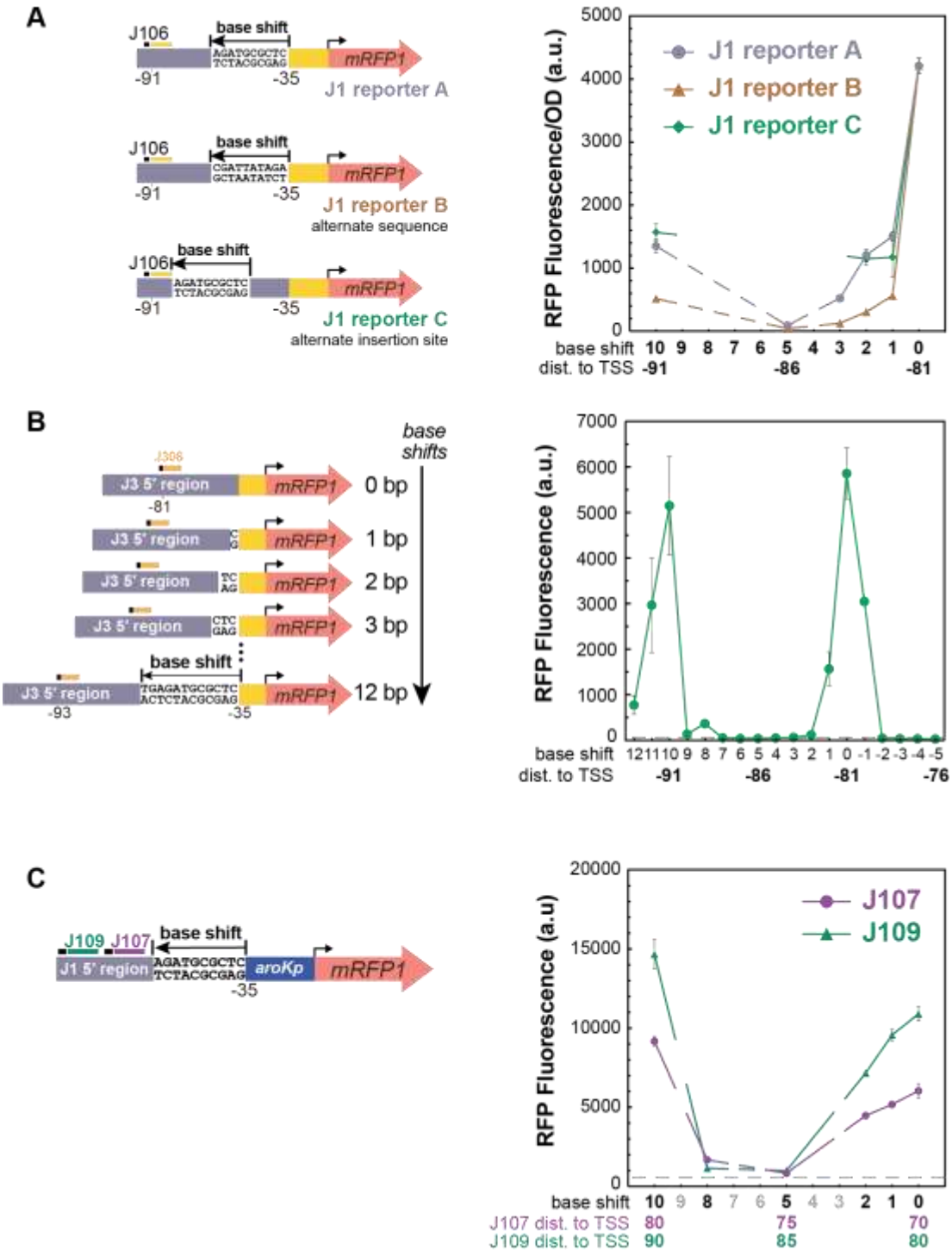
Figure S4

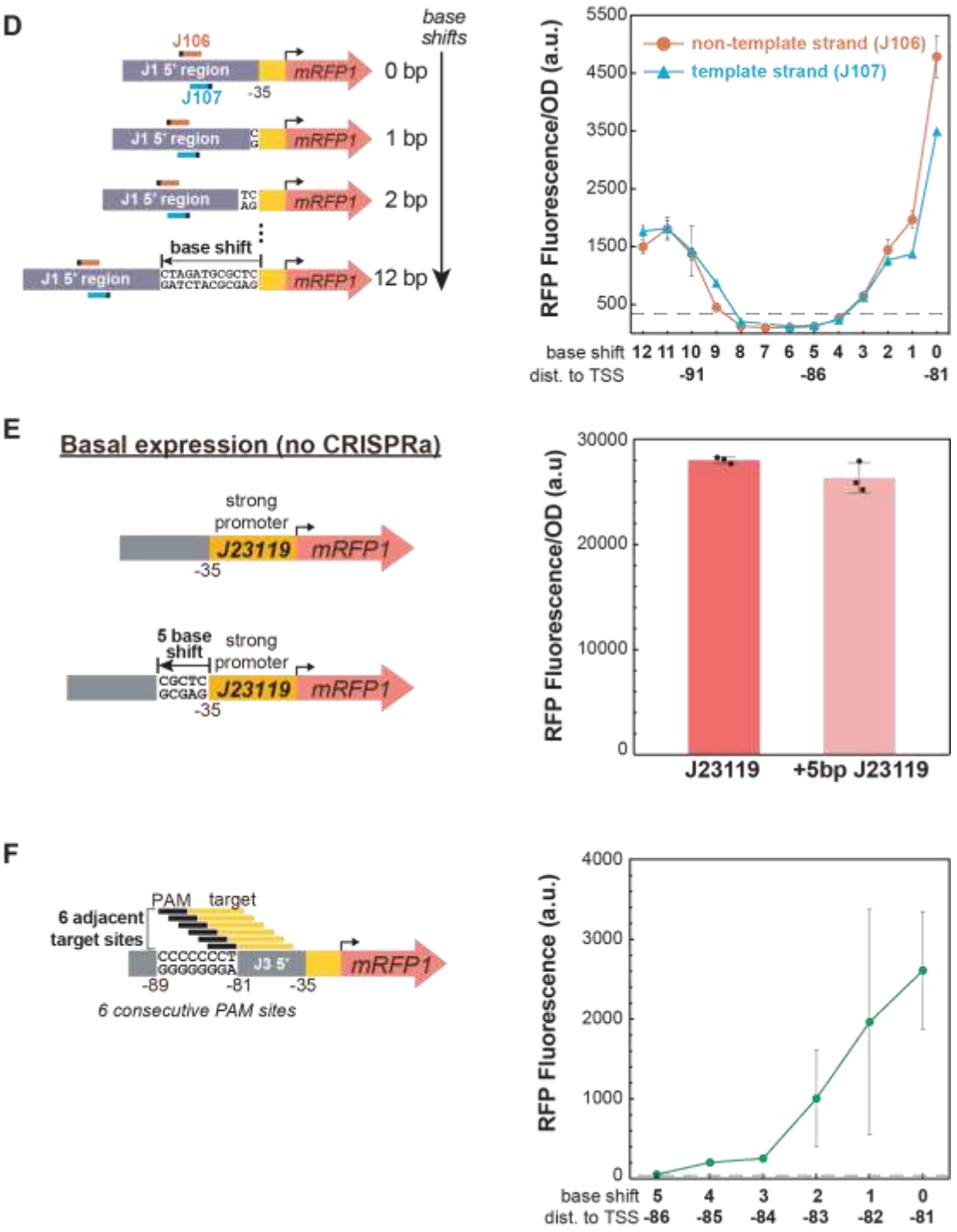


Supplementary Figure 4: CRISPRa activity depends on the target sequence on the scRNA.

Reporter cassettes that differ only by the sequence of the 20 base scRNA target site give a broad range of gene expression levels, demonstrating that the sequence of the scRNA target site can have a substantial effect on CRISPRa. Three new reporter plasmids were constructed where the J306 target site, located at -81 from the TSS, on the J3-J23117-mRFP1 reporter was replaced by the J104, J106, and J108 sequence. Activation at each promoter was tested when CRISPRa was targeted to their cognate scRNA site. The off-target negative control (OT) represents a strain expressing the original reporter with the J306 site and the CRISPRa components to target an off-target site (J206). Values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data are provided as a Source Data file.

Figure S5


A
J106
base shift
AGATGCGCTC
TCTACGCGAG
mRFP1
-91
-35
J1 reporter A
CGATTATAGA
GCTAATATCT
J1 reporter B
alternate sequence
J1 reporter C
alternate insertion site
RFP Fluorescence/OD (a.u.)
base shift
dist. to TSS
B
base shifts
J3 5' region
0 bp
1 bp
2 bp
3 bp
12 bp
TGAGATGCGCTC
ACTCTACGCGAG
-93
RFP Fluorescence (a.u.)
C
J109 J107
J1 5' region
aroKp
J107
J109
J107 dist. to TSS
J109 dist. to TSS



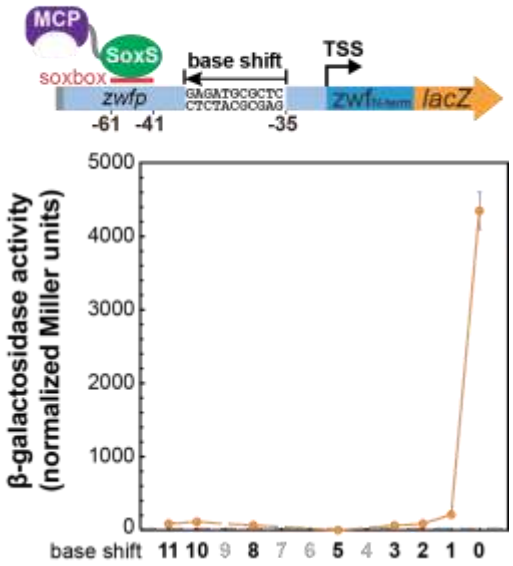
Supplementary Figure 5: The sharp positioning dependence of CRISPRa is observed across multiple promoters.

A) The sharp positioning requirements of CRISPRa are not significantly affected by the location or composition of the inserted sequence. Reporters were based on the J1-J23117-mRFP1 (Figure 4) with base shifts introduced in different ways. In the J1 reporter A, bases were inserted upstream of the -35 region. In the J1 reporter B, a different sequence was inserted at the same site. In the J1 reporter C, bases were inserted

downstream of the J106 target site. There were modest differences between reporter A and reporter B that could indicate a contribution from sequence composition. **B)** The sharp positioning requirements of CRISPRa were observed with a different heterologous promoter. CRISPRa was targeted at the -81 site on the J3-J23117 promoter that has a different upstream sequence and a different scRNA target site (J306) compared to J1-J23117. Peaks in gene expression were observed at -81 and -91 to the TSS, displaying a 10 bp periodicity. The peaks of gene expression on the J3-J23117 promoter were sharper than in the J1-J23117 promoter (Figure 4A); gene expression decreased to baseline levels after shifting only 2 bp from the peak position. After a complete 10 bp-shift period, CRISPRa activity was fully restored to the original peak expression. Reporter gene sets were constructed by inserting 0-12 bp or deleting 1-5 bp upstream of the -35 of the J3-J23117-mRFP1 reporter. The grey line represents the baseline activity of the J3-J23117-mRFP1 reporter strain containing an empty vector instead of the CRISPRa component plasmid. For comparison, previous CRISPRa data at -81 and -91 are shown on the schematic above the plot[3]. **C)** The sharp positioning requirements of CRISPRa are observed when targeting a different minimal promoter. The J1-*aroKp2* promoter displayed positioning requirements similar to the J1-J23117 promoter. Decreases in CRISPRa activity after 3 bp shifts and recovery at a 10 bp shift were observed on both the J107 and the J109 target sites. The J1-*aroKp2* promoter was constructed by replacing the BBa_J23117 minimal promoter from the J1-J23117 promoter to the aroKp2 minimal promoter. A J1-*aroKp2*-mRFP1 reporter series was constructed by adding 0 bp, 1 bp, 2 bp, 5 bp, 8 bp, and 10 bp upstream of the -35 region. The grey line represents the baseline activity of the J1-*aroKp2*-mRFP1 reporter reporter strain expressing a CRISPRa component plasmid with an off-target scRNA. **D)** The sharp positioning requirements of CRISPRa are observed on both the template and the non-template strand. For both the J106 target site on the non-template strand and the J107 target on the template strand, the CRISPRa activity decreased to baseline levels after shifting 3 bp from its original position and recovered after shifting 10 bp. The reporter plasmid was the same as in Figure 4A. The dotted grey line indicates the negative control where an off-target scRNA (J206) was co-transformed with the original reporter with no inserted bases. **E)** Adding 5 bases adjacent to the -35 region does not dramatically alter the expression of the promoter. The 5 bases added were the same as those used to shift the J1 and J3 promoters (Figure 4A and Supplementary Figure 5B). This experiment was performed using a strong minimal promoter (BBa_J23119), so that any detrimental effects would be detectable. **F)** The sharp positioning requirements of CRISPRa were observed when tested in a single reporter with multiple consecutive PAM sites. 6 consecutive PAM sites on the non-template strand were introduced by placing a CCCCCCCT sequence between -89 and -81 bp to the TSS on the J3-J23117-mRFP1 reporter. Maximum gene expression was observed at the original -81 site, after which expression gradually decreased to one third of the maximum activity after moving 2 bp away (-83). Gene expression decreased further when the scRNA target was moved 3 bp and 4 bp away from the TSS (-84 and -85), and reached the baseline

when moved 5 bp (-86). The grey line represents the baseline activity of the reporter strain containing an empty vector instead of the CRISPRa component plasmid. Values in panels A-F represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data for panels A, B, C, D, E and F are provided as a Source Data file.

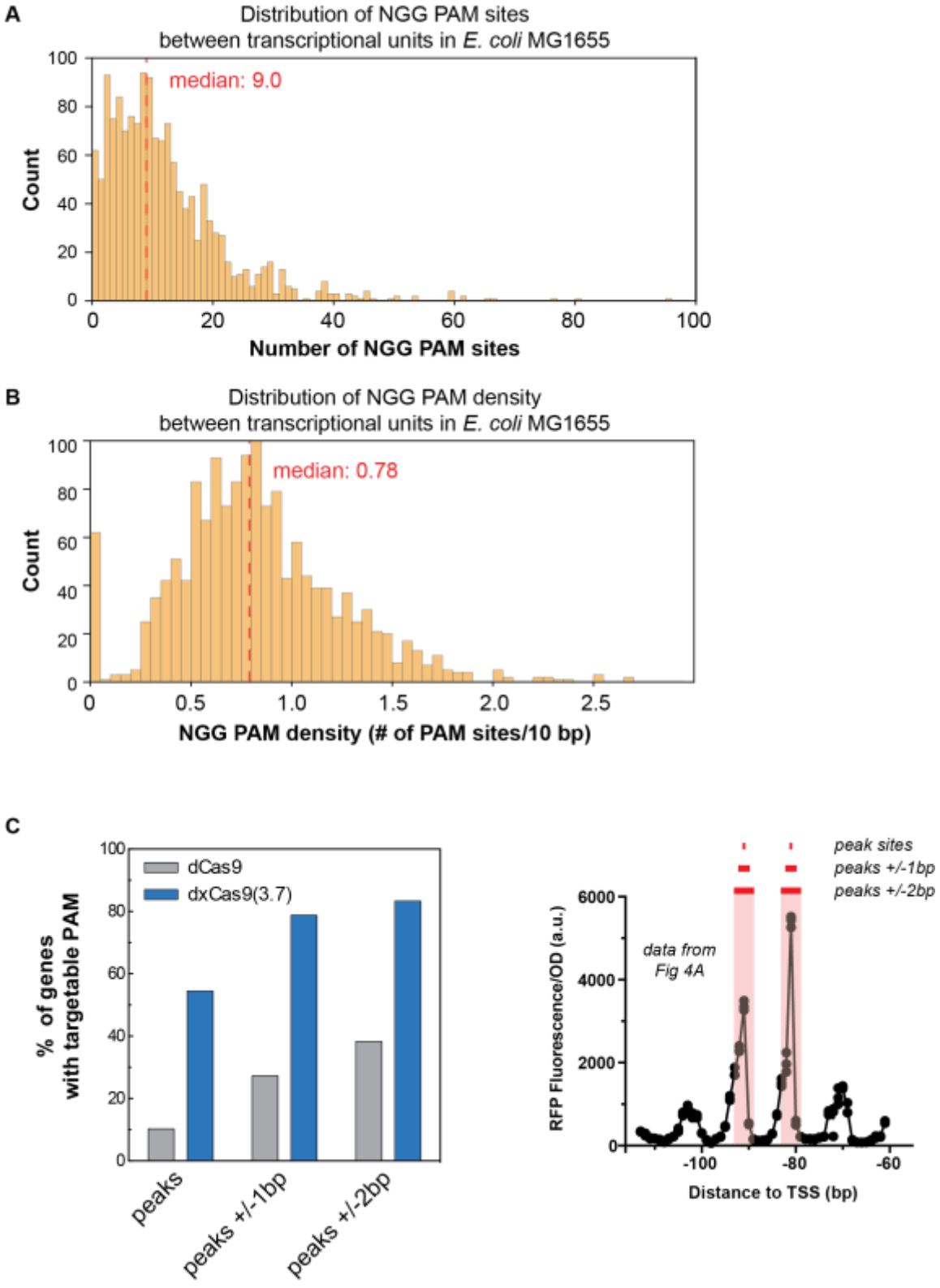
Figure S6



Supplementary Figure 6: Wild type SoxS displays a sharp positioning requirement when targeting a SoxS-dependent promoter.

When the wild type SoxS binding site (soxbox) on the endogenous *zwfp* promoter is shifted by 1 bp, gene expression decreases significantly, consistent with previous reports[4]. Shifting the soxbox further upstream causes gene expression to completely reduce to the baseline. Reporter plasmids were constructed by adding 1, 2, 3, 5, 8, 10, and 11 bp upstream of the -35 on the *zwfp-lacZ* reporter (Figure 1). Values represent the average β-galactosidase activity in normalized Miller +/- standard deviation calculated from n = 3 biologically independent samples. Source data are provided as a Source Data file.

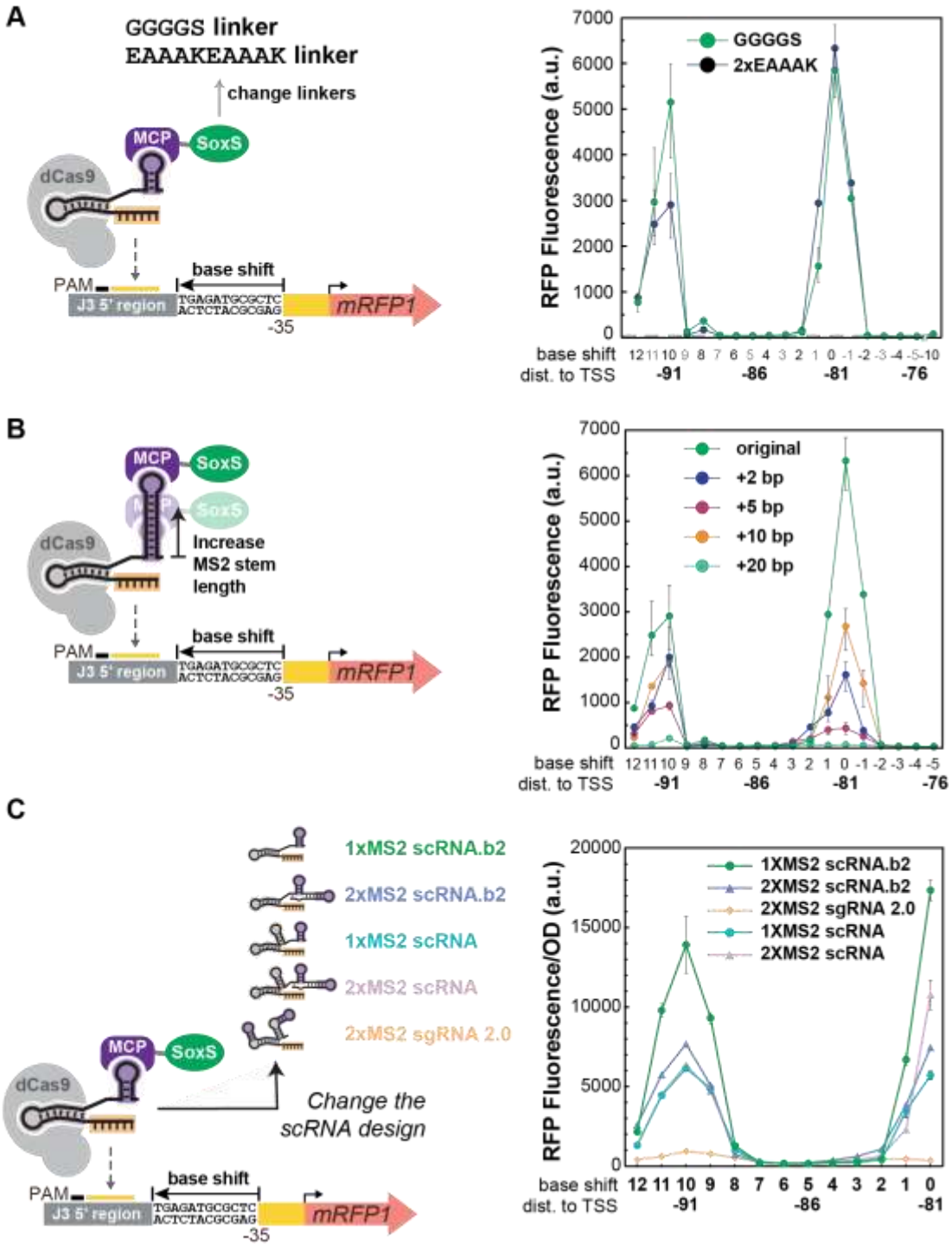
Figure S7



Supplementary Figure 7: Availability of PAM sites between transcriptional units in *E. coli* MG1655. **A)** Distribution of the number of NGG PAM sites between transcriptional units in *E. coli* MG1655. The

median number of NGG PAM sites between transcriptional units in *E. coli* MG1655 is 9.0. Methods for extracting the sequence of the DNA regions between transcriptional units in *E. coli* MG1655 and identifying available PAM sites are described in the Supplementary Methods.
B) Distribution of the density of NGG PAM sites between transcriptional units in *E. coli* MG1655. The density of PAM sites is reported over a 10 bp window. This window size was chosen because it corresponds to a full turn of the DNA helix. The density of PAM sites in a 10 bp window was calculated for each sequence between transcriptional units as the total number of PAM sites in the sequence divided by the length of the intergenic sequence and then multiplied by 10. The median density of NGG PAM sites per 10 bp between transcriptional units in *E. coli* MG1655 is 0.78. **C)** The number of genes in the *E.coli* genome that can be targeted by dCas9 is small, and the expanded PAM variant dxCas9(3.7) significantly increases the number of genes with predicted effective target sites for CRISPRa. The plot to the left shows percentage of genes with at least one PAM site targetable by dCas9 or dxCas9(3.7) at: the positions where CRISPRa displays a peak in activity. The plot to the right illustrates the range of positions on the non-template strand chosen for the analysis. Corresponding peaks on the template strand were also included. Analyses were performed using data generated when selecting candidate endogenous genes for activation (Supplementary Methods). Source data for panels A, B and C are provided as a Source Data file.

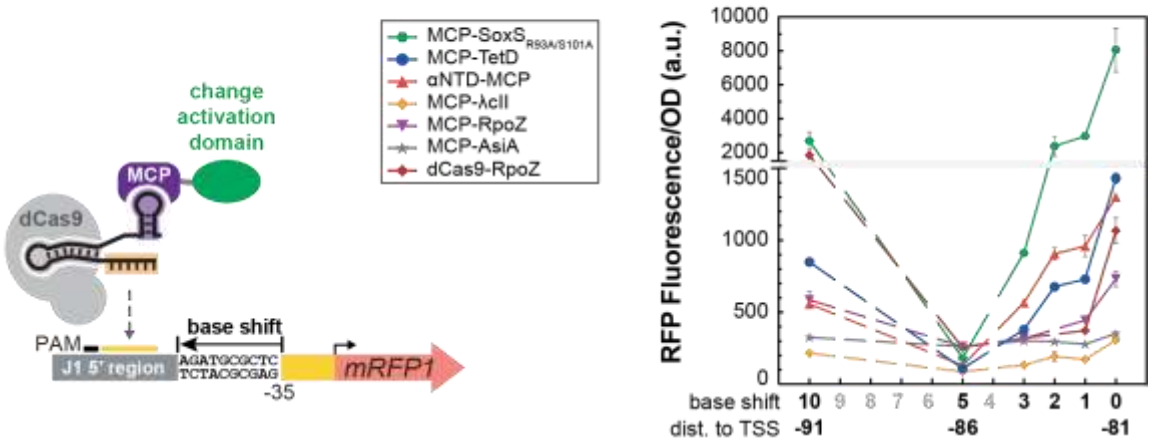
Figure S8


A
GGGGS linker
EAAAKEAAAK linker
change linkers
MCP
SoxS
dCas9
PAM
base shift
J3 5' region
TGAGATGCGCTC
ACTCTACGCGAG
mRFP1
-35
RFP Fluorescence (a.u.)
GGGGS
2xEAAAK
base shift
dist. to TSS
-91
-86
-81
-76
B
Increase
MS2 stem
length
original
+2 bp
+5 bp
+10 bp
+20 bp
C
1xMS2 scRNA.b2
2xMS2 scRNA.b2
1xMS2 scRNA
2xMS2 scRNA
2xMS2 sgRNA 2.0
Change the
scRNA design
RFP Fluorescence/OD (a.u.)
1XMS2 scRNA.b2
2XMS2 scRNA.b2
2XMS2 sgRNA 2.0
1XMS2 scRNA
2XMS2 scRNA

Supplementary Figure 8: Modifying the CRISPRa complex structure does relax the sharp positioning requirements of CRISPRa.

A) Changing the linker between MCP and SoxS does not change the positioning dependence of CRISPRa. A CRISPRa complex with the MCP-(EAAAKEAAAK)-SoxS(R93A/S101A) activation domain displayed 10 bp periodicity of peak expression similar to that with the MCP-(GGGGS)-SoxS(R93A/S101A) activation domain. The EAAAKEAAAK linker is predicted to be more rigid than the GGGGS linker[5]. The grey line represents the baseline activity of the J3-J23117-mRFP1 reporter strain containing an empty vector instead of the CRISPRa component plasmid. **B)** Extending the length of the MS2 stem does not change the positioning dependence of CRISPRa. CRISPRa systems with scRNAs that have +2, +5, and +10 RNA base pairs added to the bottom of the MS2 stem displayed the same 10 bp periodicity but lower peak activity compared to the original 1xMS2 scRNA.b2. The strain having the scRNA with +20 bp extended MS2 stem did not show any CRISPRa activity. **C)** No improvements in the effective target range were observed for CRISPRa systems with alternative scRNA designs. The scRNA designs tested are: 1xMS2 scRNA.b2 and 2XMS2 scRNA.b2 with the tracrRNA hairpin removed[3], the original 1xMS2 and 2XMS2 scRNA design having the tracrRNA hairpin[6], and sgRNA 2.0 where the MS2 hairpins are extended from the RNA stems on the sgRNA[7]. CRISPRa systems expressing a 2XMS2 scRNA.b2, 1xMS2 scRNA, and 2XMS2 scRNA displayed the same 10 bp periodicity but lower peak activity than 1xMS2 scRNA.b2. Expressing a sgRNA 2.0 resulted in no CRISPRa activity. Values in panels A-C represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data for panels A, B and C are provided as a Source Data file.

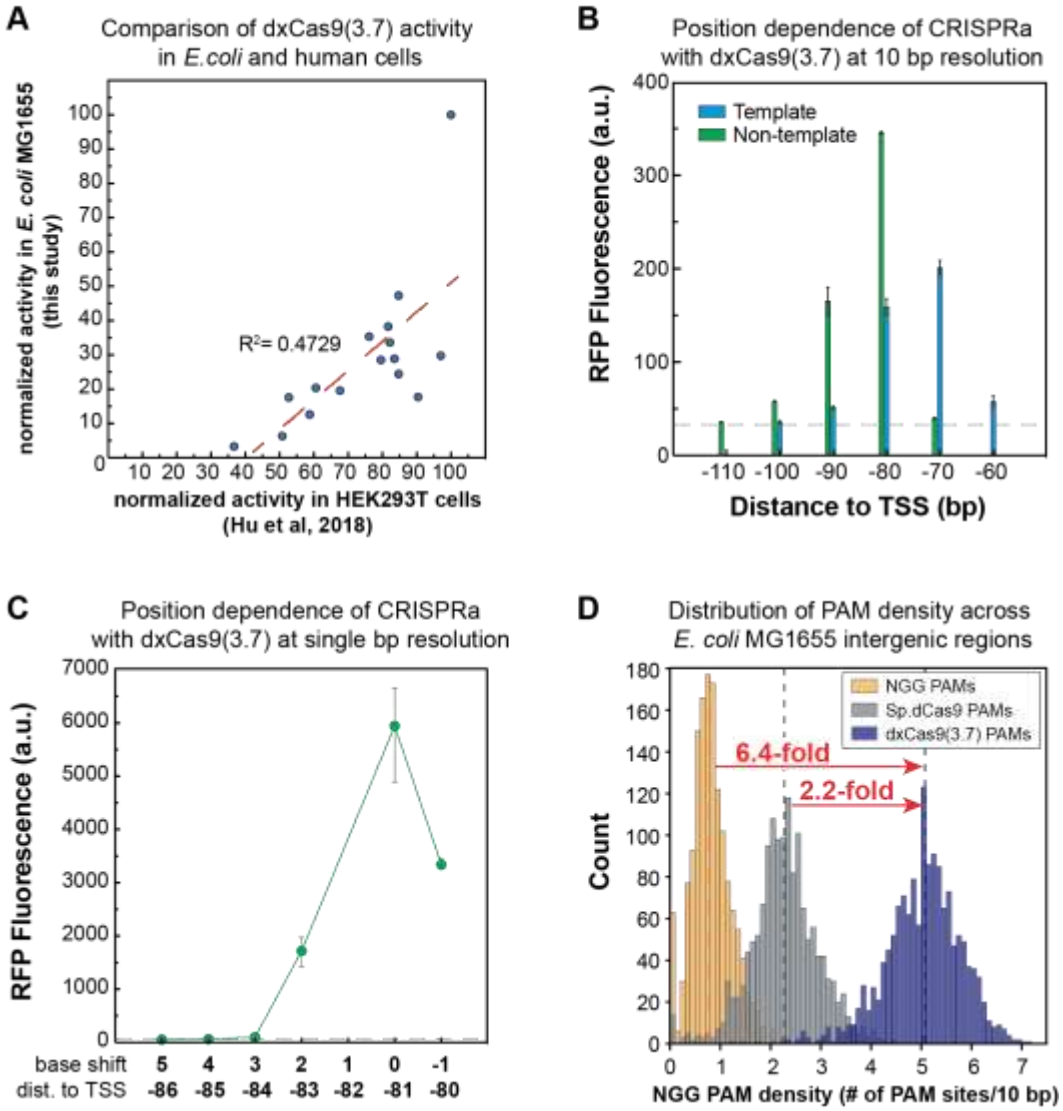
Figure S9



Supplementary Figure 9: Performing CRISPRa with alternative activation domains does not expand the range of targetable positions.

CRISPRa with MCP-TetD and αNTD-MCP activation domains[3] displayed similar positioning dependence as MCP-SoxS(R93A/S101A), where gene expression decreases as the position shifts from 0 to 3 bp, reaches the baseline at a 5 bp shift, and increases again at a 10 bp shift. CRISPRa with MCP-RpoZ[3] and dCas9-RpoZ[8] activation domains displayed more stringent positioning dependence, where gene expression approaches the baseline at 1 bp shift. All alternative activation domains gave weaker peak gene expression at 0 bp shift compared to MCP-SoxS(R93A/S101A). MCP-λcII and MCP-AsiA[3] activation domains did not show any significant CRISPRa activity. All activation domains were cloned into the CRISPRa component plasmid containing *Sp*-dCas9 and 1xMS2 scRNA.b2 targeting J106. To reduce the toxicity of AsiA, the MCP-AsiA plasmid also contains a co-expressed *rpoD(F563Y)* gene[3]. The MCP-SoxS(R93A/S101A), MCP-TetD, αNTD-MCP, MCP-λcII and MCP-AsiA plasmids were tested in *E. coli* MG1655. The MCP-*rpoZ* and dCas9-*rpoZ* plasmids were tested in the *E. coli* strain CD03 (MG1655/Δ*rpoZ*) (Supplementary Table 1)[3]. Values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. Source data are provided as a Source Data file.

Figure S10

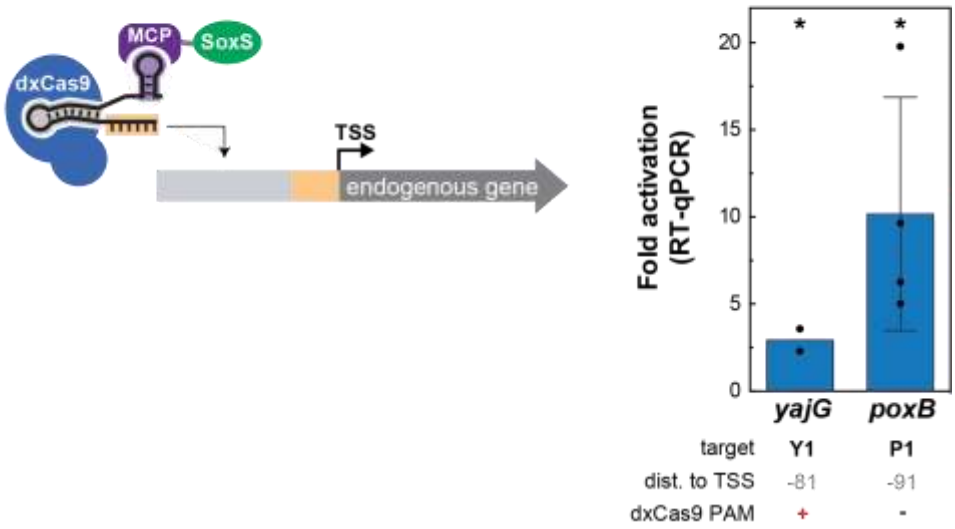


Supplementary Figure 10: dxCas9(3.7) can target an expanded range of PAM sites and is sensitive to target site position for CRISPRa.

A) Relative CRISPRa activities of dxCas9(3.7) on different PAM targets in *E. coli* correlates well with the corresponding data obtained by Hu et al.[9] in human cells. The CRISPRa activity on the NGG PAM site in both studies were normalized to 100 and the CRISPRa activity on all the other PAM sites were normalized to the value of NGG PAM site. Normalized data in this study (y axis) were plotted against the normalized data obtained by Hu et al. (x axis). **B)** CRISPRa with dxCas9(3.7) displayed similar sensitivity to target site position as *Sp*-dCas9. Peaks of activation were observed at -81 and -91 on the non-template strand and -70 and -80 on the template strand. CRISPRa was targeted to a J1-J23117-mRFP1 reporter integrated into the genome (*E. coli* CD13, Supplementary Table 1). Values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. The grey dotted line represents the baseline

fluorescence of a strain containing the dxCas9(3.7) and MCP-SoxS(R93A/S101A) and an empty vector with no scRNAs. **C)** CRISPRa with dxCas9(3.7) displayed similar positioning dependence with single base shifts. Gene expression was significantly reduced when the scRNA target site was shifted 1-2 bp in either directions from the optimal position -81 bp to the TSS on the non-template strand. Shifting the scRNA target site 3-5 bp causes gene expression to fall to the baseline level. Reporter gene sets were constructed with 1 bp deleted or 2-5 bp inserted upstream of the -35 of the J3-J23117-mRFP1 reporter. The grey line represents the baseline activity of a strain containing the J3-J23117-mRFP1 reporter plasmid and a CRISPRa component plasmid with an off-target scRNA (hAAVS1). Values represent the average +/- standard deviation calculated from n = 3 biologically independent samples. **D)** dxCas9(3.7) increases the probability of finding targetable PAM sites for CRISPRa. Histograms showing the distribution of the PAM density (defined in Supplementary Figure 7B) between transcriptional units in *E. coli* MG1655 suggest that the average likelihood of finding a dxCas9(3.7)-compatible PAM (blue) is ~6.4-times higher than finding an NGG PAM (yellow), and ~2.2-times higher than finding a *Sp*-dCas9-compatible PAM (grey). Vertical dotted lines indicate the median PAM density for each group. The median PAM density is 5.08 for dxCas9(3.7)-compatible PAMs (NGG, AGA, AGC, AGT, CGA, CGC, CGT, GGA, GGC, GGT, TGA, TGC, TGT, GAA, GAT, CAA), 2.33 for *Sp*-dCas9-compatible PAMs (NGG, AGA, CGA, GGA, GGC, GGT, TGA). and 0.78 for NGG PAMs. Methods for extracting the sequence of *E. coli* MG1655 intergenic regions, counting targetable PAM sites and calculating the PAM density are described in the Supplementary Methods. Source data for panels A, B, C and D are provided as a Source Data file.

Figure S11



Supplementary Figure 11. Predictive rules for CRISPRa enable activation of endogenous genes *yajG* and *poxB*.

The scRNA sites displaying the highest activity on the *yajG* and *poxB* reporters from the *E. coli* promoter collection[10] (Figure 6B) were tested for activity with dxCas9(3.7) at the endogenous *yajG* and *poxB* genes using RT-qPCR. Fold activation represents expression levels relative to a control expressing an off-target scRNA (J306). For Y1, the value represents the average calculated from n = 2 biologically independent samples. For P1, the value represents the average +/- standard deviation calculated from n = 3 biologically independent samples. The (*) symbol indicates a statistically significant difference from the off-target control (*p-value* < 0.05 using a two-tailed unpaired Welch's t-test). Exact *p-values*: Y1: 0.02, P1: 0.04. Source data are provided as a Source Data file.

Supplementary tables

Supplementary Table 1. *E. coli* Strains

Strain	**Description**	**Genotype**	**Reference**
MG1655	parent *E. coli* strain	F- λ- ilvG- rfb-50 rph-1	
CD03	MG1655 with *rpoZ* knocked out	MG1655 Δ*rpoZ*	[3]
CD06	MG1655/sfGFP (weak promoter)	MG1655 *W1-BBa_J23117-sfGFP KanR::nfsA*	[3]
CD13	MG1655/mRFP1 (weak promoter)	MG1655 *J1-BBa_J23117-mRFP1::nfsA*	This study, Suppl. Fig. 8

Supplementary Table 2. Description of the CRISPRa systems used in each figure

Figure	Cas protein	scRNA design	scRNA target	Activation domain
1B	dCas9	1xMS2 scRNA.b1	W108	MCP-(5aa)-SoxS(wild-type or mutant)
2A	dCas9	1xMS2 scRNA.b1	A1-A4, hAAVS1	MCP-(5aa)-SoxS(R93A/S101A)
2B	dCas9	1xMS2 scRNA.b2	C1-C5, hAAVS1	MCP-(5aa)-SoxS(R93A)
3A	dCas9	1xMS2 scRNA.b2	J306, J206	MCP-(5aa)-SoxS(R93A)
3B	dCas9	1xMS2 scRNA.b2	J101-J120, hAAVS1	MCP-(5aa)-SoxS(R93A/S101A)
3C	dCas9	1xMS2 scRNA.b2	J306, J206	MCP-(5aa)-SoxS(R93A)
3D	dCas9	1xMS2 scRNA.b2	J306, J206	MCP-(5aa)-SoxS(R93A/S101A)
4A	dCas9	1xMS2 scRNA.b2	J102, J104, J106, J108, J110, J206	MCP-(5aa)-SoxS(R93A)
4B	dCas9	1xMS2 scRNA.b2	J106	MCP-(5aa)-SoxS(R93A), MCP-(10aa)-SoxS(R93A), MCP-(20aa)-SoxS(R93A)
5A	dCas9 / dxCas9(3.7)	1xMS2 scRNA.b2	J306	MCP-(5aa)-SoxS(R93A/S101A)
5B	dCas9 / dxCas9(3.7)	1xMS2 scRNA.b2	M1, M2, J206	MCP-(5aa)-SoxS(R93A/S101A)
6A	dCas9 / dxCas9(3.7)	1xMS2 scRNA.b2	Y1-Y3, J306	MCP-(5aa)-SoxS(R93A/S101A), tet-inducible
6B	dxCas9(3.7)	1xMS2 scRNA.b2	Y1, Y2, P1, P2, U1, U2, D1, D2, B1, B2, E1, E2, J306	MCP-(5aa)-SoxS(R93A/S101A), tet-inducible

Supplementary Figure 1	dCas9	1xMS2 scRNA.b2	L1-L8, hAAVS1	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 3A	dCas9	1xMS2 scRNA.b2	J306, J206	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 4	dCas9	1xMS2 scRNA.b2	J104, J106, J108, J306, J206	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 5A	dCas9	1xMS2 scRNA.b2	J106	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 5B	dCas9	1xMS2 scRNA.b2	J306	MCP-(5aa)-SoxS(R93A/S101A)
Supplementary Figure 5C	dCas9	1xMS2 scRNA.b2	J107, J109	MCP-(5aa)-SoxS(R93A/S101A)
Supplementary Figure 5D	dCas9	1xMS2 scRNA.b2	J106, J107, J206	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 5F	dCas9	1xMS2 scRNA.b2	J306, J306+1-5	MCP-(5aa)-SoxS(R93A/S101A)
Supplementary Figure 8A	dCas9	1xMS2 scRNA.b2	J306	MCP-(5aa)-SoxS(R93A/S101A), MCP-(2xEAAAK)-SoxS(R93A/S101A)
Supplementary Figure 8B	dCas9	1xMS2 scRNA.b2, 1xMS2 scRNA.b2 + 2/5/10 bp MS2 stem extension	J306	MCP-(2xEAAAK)-SoxS(R93A/S101A)
Supplementary Figure 8C	dCas9	1xMS2 scRNA.b2, 2xMS2 scRNA.b2, 2xMS2 sgRNA2.0, 1xMS2 scRNA, 2xMS2 scRNA	J306	MCP-(5aa)-SoxS(R93A)

Supplementary Figure 9	dCas9	1xMS2 scRNA.b2	J106	MCP-(5aa)-SoxS(R93A/S101A), Alternative activators
Supplementary Figure 10B	dxCas9(3.7)	1xMS2 scRNA.b2	J104-J113	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 10C	dxCas9(3.7)	1xMS2 scRNA.b2	J306	MCP-(5aa)-SoxS(R93A)
Supplementary Figure 11	dxCas9(3.7)	1xMS2 scRNA.b2	Y1, P1, J306	MCP-(5aa)-SoxS(R93A/S101A), tet-inducible

Supplementary Table 3. gRNA Target Sites

sgRNA target	DNA Sequence	Target Strand[a]	Distance to TSS[b]
W108	GAAGATCCGGCCTGCAGCCA	NT	91
J101[c]	TGGGTTCCACCGGATACCTC	T	40
J103[c]	AGGCGTCCTTTGGGTTCCAC	T	50
J105[c]	CGGTTACCAAAGGCGTCCTT	T	60
J107[c]	CGGTGTCCTGCGGTTACCAA	T	70
J109[c]	AGGTATCCTGCGGTGTCCTG	T	80
J111[c]	GGGCGACCTCAGGTATCCTG	T	90
J113[c]	GGGCCACCACGGGCGACCTC	T	100
J115[c]	TGGTGACCATGGGCCACCAC	T	110
J117[c]	GGGTGACCTATGGTGACCAT	T	120
J119[c]	TGGTTGCCAAGGGTGACCTA	T	130
J121[c]	AGGACACCTTTGGTTGCCAA	T	140
J102[c]	AGGTATCCGGTGGAACCCAA	NT	61
J104[c]	TGGAACCCAAAGGACGCCTT	NT	71
J106[c]	AGGACGCCTTTGGTAACCGC	NT	81
J108[c]	TGGTAACCGCAGGACACCGC	NT	91
J110[c]	AGGACACCGCAGGATACCTG	NT	101
J112[c]	AGGATACCTGAGGTCGCCCG	NT	111
J114[c]	AGGTCGCCCGTGGTGGCCCA	NT	121
J116[c]	TGGTGGCCCATGGTCACCAT	NT	131
J118[c]	TGGTCACCATAGGTCACCCT	NT	141
J120[c]	AGGTCACCCTTGGCAACCAA	NT	151
hAAVS1[c]	GGGGCCACTAGGGACAGGAT	off-target	n/a
J206	TAGTAGCCGAACACGTCCTC	off-target	n/a
J306	TTGTGTCCAGAACGCTCCGT	NT	81
J306+1	TGTGTCCAGAACGCTCCGTA	NT	82
J306+2	GTGTCCAGAACGCTCCGTAG	NT	83
J306+3	TGTCCAGAACGCTCCGTAGG	NT	84
J306+4	GTCCAGAACGCTCCGTAGGG	NT	85
J306+5	TCCAGAACGCTCCGTAGGGG	NT	86

M1	AGCAGAAGTGTCAGCAGTGT	NT	81 (reporter A), 86 (reporter B)
M2	CGACGAGCAGAAGTGTCAGC	NT	76 (reporter A), 81 (reporter B)
aroKB_A1	GGGCAATTATTTCGTCATGA	T	151
aroKB_A2	AGATGAACGACGCGAGTTAG	T	122
aroKB_A3	TTTTACGGCTGTTTACTCAC	NT	92
aroKB_A4	TGAGTAAACAGCCGTAAAAG	T	71
cysK_C1	CCACCCCTGTTTCACACAAA	NT	93
cysK_C2	AAACCGTTTGTGTGAAACAG	T	79
cysK_C3	GACATGCAAGATGGAATAAG	NT	61
cysK_C4	ATGACATGCAAGATGGAATA	NT	59
cysK_C5	GGAAATAATGACATGCAAGA	NT	52
ldhA_L1	CAGTAATAACAGCGCGAGAA	T	157
ldhA_L2	GGATATTAACTACCCATGCT	NT	140
ldhA_L3	GCTTTATATTTACCCAGCAT	T	135
ldhA_L4	GCTTAATTTTTCGCTAAATC	NT	119
ldhA_L5	GAAAAATTAAGCATTCAATA	T	91
ldhA_L6	AGCATTCAATACGGGTATTG	T	82
ldhA_L7	AGGCGCAACCTTCAACTGAA	NT	67
ldhA_L8	ATGTTTAACCGTTCAGTTGA	T	58
yajG_Y1	TTGACGAAATAATCGCCCCT	NT	81
yajG_Y2	CATCAGTGTTTCTTTTACCA	T	80
yajG_Y3	AAATAATCGCCCCTGGTAAA	NT	87
poxB_P1	CCCGATGAAAGGAATATCAT	NT	91
poxB_P2	GGTTAAATAGCCCGATGAAA	NT	81
uxuR_U1	TGATTGACCAGTAAGTCTGT	NT	81
uxuR_U2	GATTACCCTACAGACTTACT	T	70
ppiD_D1	ACTAAGCGTTGTCCCCAGTG	T	80
ppiD_D2	GTCCCCAGTGGGGATGTGAC	T	70
ansB_B1	AGATCTACAAAGTTAGAGGC	NT	91
ansB_B2	TATATTTTGGAGATCTACAA	NT	81
araE_E1	TGCGACATGTCGTTATGTGA	NT	91

araE_E2	ATTAAATTGCTGCGACATGT	NT	81

[a] Template strand (T) or non-template strand (NT).

[b] Distance to TSS is the distance from the 3' end (PAM proximal) of the guide target site to the transcription start site. For synthetic promoters driven by BBa_J23117 or BBa_J23119 (http://parts.igem.org), the TSS is immediately downstream of the BBa sequence (see complete maps below).

[c] The J101-J121 sites were the same target sites used to test the positioning dependence of CRISPRa at a 10 bp resolution on the J1-J23117 promoter[3].

Supplementary Table 4. Select *E. coli* Expression Plasmids[a]

Plasmid	Marker	origin	Promoter	Gene	Terminator
pCD442	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107*	1) dCas9 2) MCP-(5aa)-SoxS (R93A/S101A)	*1) BBa_B0015* *2) BBa_B1002*
pCK005.1-21[b]	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107* *3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R93A/S101A) 3) 1x MS2 scRNA.b2 (J101-121 targets)	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*
pCD564	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107*	1) dxCas9(3.7) 2) MCP-(5aa)-SoxS (R93A/S101A)	*1) BBa_B0015* *2) BBa_B1002*
pCD565	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107* *3) BBa_J23119*	1) dxCas9(3.7) 2) MCP-(5aa)-SoxS (R93A/S101A) 3) 1x MS2 scRNA.b2 (J306 target)	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*
pCD580.-1~-5	*AmpR*	*pSC101***	*J3_BBa_J23117 (with 1-5 bp deleted from the J1 region)*	mRFP1	*BBa_B0015*
pCD581	*CmR*	*p15A*	*1) Sp.pCas9* *2) BBa_J23107* *3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R93A/S101A) 3) 1x MS2 scRNA.b2 (J306 target)	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*
pJF215.x	*CmR*	*p15A*	*1) Sp.pCas9* *2) TetR-pTet* *3) BBa_J23119*	1) dCas9 2) MCP-(5aa)-SoxS (R93A/S101A) 3) 1x MS2 scRNA.b2 (variable target)	*1) BBa_B0015* *2) BBa_B1002* *3) TrrnB*
pJF076Sa[c]	*AmpR*	*pSC101***	*J3_BBa_J23117*	mRFP1	*BBa_B0015*

pJF143-J3[d]	*AmpR*	*pSC101***	*J3_BBa_J23117*	mRFP1	*BBa_B0015*
pJF155.1-12[c]	*AmpR*	*pSC101***	*J1_BBa_J23117 (with 1-12 bp inserted upstream of -35)*	mRFP1	*BBa_B0015*
pJF161.1-12[d]	*AmpR*	*pSC101***	*J3_BBa_J23117 (with 1-12 bp inserted upstream of -35)*	mRFP1	*BBa_B0015*

[a] BBa sequences are from the Repository of Standard Biological Parts (http://parts.igem.org). dCas9 is the catalytically inactive form of *S. pyogenes* Cas9. Sp.pCas9 is the endogenous Cas9 promoter from *S. pyogenes*.

[b] pCK005.1-21 indicates a set of plasmids (pCK005.1, pCK005.2…) where the final number corresponds to guide RNA target sites (J101, J102…, Supplementary Table 3) used for the J1-117-mRFP1 reporter.

[c] Originally described in previous work[3]. Modified versions of this plasmid are available where BBa_J23117 is replaced with minimal promoters regulated by alternative sigma factors (Figure 3B).

[d] Modified version of this plasmid are available where: BBa_J23117 is replaced with Anderson promoters of different strength (Figure 3A) and different PAM sites at the -81 J306 site (Figure 5A).

Supplementary Table 5. Primer Sequences for RT-qPCR

Primer	Sequence	Reference
16S_f	AAAGTTAATACCTTTGCTCATTGACGTT	[3]
16S_r	GACTACCAGGGTATCTAATCCTGTTT	[3]
aroK_f	TCTGGTTGGGCCTATGGGTG	This study
aroK_r	TACGAATGGTCACGTCGGCA	This study
cysK_f	TGCTGAAACCAGGCGTTGAA	This study
cysK_r	TCCCAACGCCAGCAATAAATACA	This study
ldhA_f	TGGCTGCGAAGCGGTATGTA	This study
ldhA_r	GAACGCCAGCAGACGCATAC	This study
yajG_fw	AAGTCACCCGCGATAAT	This study
yajG_rev	CTTTGGTCGCGATGTTG	This study
poxB_fw	GGTCTTAGTGACAGTCTTAATC	This study
poxB_r	GGAATATGAGCGGCAATC	This study

Supplementary Table 6. Intervening sequences between scRNA target site and -35 region with respective CRISPRa activity and number of transcription factor (TF) binding sites.

Index	sfGFP/OD_{600}[a]	Sequence[b]	TF binding sites, consensus[c]	TF binding sites, *d(u,v)* ≤ 1[c]
1	549.87	TATCATAGTGATGACCAGTAAACATT	0	4
2	602.70	TGCAGAAAAGGGCTCTAGTGACTTGT	0	7
3	682.05	ATCTATGCGACGTCAAACGTGATGGG	0	12
4	702.47	GCCTTTCAGTGATCCCTTGTGATCTT	0	11
5	704.82	CAATTCTAGCAAGAAATTGACTTCGT	0	6
6	710.25	TTTACGGATTGGTGCGCTATTGGGTT	0	14
7	716.04	TAATGTTGATCGCCTCTATGCGTCCT	0	5
8	746.08	TTATTATAATAGTTTTGAACGTGCCT	0	6
9	794.37	CAATATGACGTGTTGTTAATTTGGTT	1	14
10	883.74	GGTCTGTGTACTCGAACACGAGATTT	0	7
11	1230.90	TTGCTCCGCGGTTTCTCTGTAAATGT	0	8
12	1532.80	ATAAATTGCATGATTAAAGACTATTG	0	11
13	1586.70	TTCGGATTTAAGGAGATATGTTTACG	0	7
14	1679.31	TAATAGGCTAGGAATTTGAAGGGATT	0	11
15	5604.77	GCATCTACATAGTGGTACAATTGAAG	0	13
16	7110.46	GCGTCCATATATGCTGATGGTGTGGG	0	10
17	7143.40	GCGGGCTATGCCTTGGAGGGTTGTAG	0	12
18	11027.55	ATTACACACGAAGGTGTATTTACTAT	2	28
19	11589.46	TTATCTAATAGGGGCGCCCCCGCTTC	0	9
20	11714.54	TTAAACAAACGAAAAGTTCGCGCCTG	1	14
21	12073.23	GATCGGGTAGGGCGCCGACAAAGGGA	0	8
22	12374.02	TGAGGGATATAGAGAGATGTTCAGCC	0	11
23	12399.33	CATCGCCGTTGCGATGCATTTTGACG	0	15
24	13207.42	GTTGTCCTTCTAGTCGCCCATGACTC	0	8
J3	**13269.13**	**CGTCGTCTTGAAGTTGCGATTATAGA**	0	10
25	13413.88	GTCGTAAATAAGTAAGTCACTCCCAC	2	19
26	13434.32	TTGAGGCCCATGCTTGTGGAAATGAC	1	20
27	14053.61	GGCAAGATGCCTCGTGCAGTAGAATA	0	8
28	15973.97	AGTCGCTGAGTAGATGTTCGTAGAGA	0	4
29	16760.50	ACACCGACTACCCCTGCTGGGCCCAG	0	12

[a] sfGFP/OD_{600} values represent the CRISPRa activity of strains where CRISPRa is targeted to individual elements of a reporter library where the 26 bases between the scRNA target site and the -35 region on the

J3-J23117-sfGFP reporter were replaced with random bases. The sfGFP/OD_{600} of a strain expressing the J3-J23117-sfGFP promoter are included for comparison and are in bold.

[b] The sequence corresponds to the 26 bases between the scRNA target site and the -35 region of the promoter depicted in Figure 3C.

[c] Number of exact matches to consensus transcription factor binding sites[2] or sequences within a Hamming distance $d(u,v) \leq 1$ from the consensus transcription factor binding sites.

Supplementary methods

Computational analysis of PAM site availability

The intergenic sequences from *E. coli* were obtained from the RegulonDB database[2]. To obtain the intergenic sequences upstream of promoters, the intergenic sequences between convergent genes and between intra-operon coding sequences were removed. 5' untranslated regions (UTRs) were removed from sequences using transcription start site (TSS) information from Kim et al.[11] and RegulonDB. Intergenic sequences upstream of genes with no known TSS or where no intergenic sequence remained after removing the UTRs were discarded, yielding 1504 transcriptional units for further analysis. The number of PAM sites was calculated by counting the number of PAM sites on both strands (Supplementary Figure 10D). The PAM density over a 10 bp window was calculated by dividing the number of PAMs found at each intergenic sequence by the length of the intergenic sequence and multiplying by 10. Analyses were performed using Python 2.7.

Computational selection of candidate endogenous genes for CRISPRa

PAM sites found in *E. coli* intergenic sequences upstream of promoters were assigned a "sequence score" and a "distance score". The sequence score indicates the relative activity of a given PAM sequence compared to an AGG PAM, and was calculated using data from Figure 5A with dxCas9(3.7) (e.g. AGG = 1, CGT = 0.24). All NGG sequences were assigned a "sequence score" of 1. The distance score indicated the relative activity of a given PAM site position compared to a PAM site found at -81 on the non-template strand or -70 on the template from the TSS. Distance scores for the non-template strand were calculated using data from Figure 4A (e.g. -81 = 1, -91 = 0.62). For the template strand, the same scores were used, but each score was shifted upstream by 11 bases to match the position of the site of maximum activation (e.g. -70 = 1, -80 = 0.62). Each PAM site was assigned a final calculated score as "sequence score" x "position score". Promoters were then ranked by sorting for higher PAM scores at the peaks of activation (sum of the scores for the PAMs at -70, -80, -81, -91). This analysis was performed using Python 2.7. Candidates were then manually selected from among the top scoring promoters by the following criteria: (1) two or more candidate PAM sites, (2) regulation by σ^{70}, and (3) relatively weak basal expression level (<10% of the level of the maximally expressed *E. coli* gene according to data from the *E. coli* promoter collection[10]. Using these criteria, we chose the following six candidates for further characterization: *yajG*, *uxuR*, *ansB*, *poxB*, *araE*, and *ppiD*. Each of these genes is represented in the *E. coli* promoter collection (Dharmacon), a commercially available library of promoter-GFPmut2 fusions[10].

Sequence of the double mutant activation domain: MCP-(5aa)-SoxS(R93A/S101A)

> MCP-(5aa)-SoxS(R93A/S101A) (optimized for minimal endogenous activity, Figure 1 and subsequent)

$MCP_{\Delta FG, V29I}$., 5 aa linker, SoxS(R93A/S101A) (underlined are the alanine point mutations)

MGPASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKG
AWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGGSMSHQKIIQDLIAWIDE
HIDQPLNIDVVAKKSGYSKWYLQRMFRTVTHQTLGDYIRQRRLLLAAVELRTTERPIFDIAMDLGYVSQQ
TFSRVFARQFDRTPADYRHRL

Reporter genes

The integrated sfGFP reporter, the *zwfp-lacZ* and *fumCp-lacZ* reporters used for testing the CRISPRa and endogenous activities of mutant SoxS (Figure 1) and the J1-J23117-mRFP1 reporter (Figure 3 & Supplementary Figure 5 and subsequent) for testing the distance dependence property were described in previous study[3].

J3-J23117_mRFP1 reporter (Figure 3 & Supplementary Figure 4 and subsequent)

The J3 upstream sequence contains a PAM site allowing guide RNAs to target at -81 bp to the TSS on the non-template strand, which is the same distance where maximum CRISPRa activity was observed on the J1 upstream region described previously[3].

J3 upstream region, BBa_J23117 promoter, Bujard RBS, mRFP1, BBa_B0015 terminator. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGATTGACAGCTAGCTCAGTCCTAGGGATTGTGCTAGCGAATT
CATTAAAGAGGAGAAAGGTACCATGGCGAGTAGCGAAGACGTTATCAAAGAGTTCATGCGTTTCAAAGTT
CGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAGGTGAAGGTGAAGGTCGTCCGTACGAAG
GTACCCAGACCGCTAAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTTGGGACATCCTGTCCCC
GCAGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCGGACTACCTGAAACTGTCC
TTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCGAAGACGGTGGTGTTGTTACCGTTACCCAGG
ACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTGGTACCAACTTCCCGTCCGACGG
TCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCGAACGTATGTACCCGGAAGACGGTGCT
CTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACGGTGGTCACTACGACGCTGAAGTTAAAACCA
CCTACATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCTTACAAAACCGACATCAAACTGGACATCACCTC
CCACAACGAAGACTACACCATCGTTGAACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAA
GGATCCAAACTCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTT

TCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCT
TTCTGCGTTTATA

J3(N26)-J23117-sfGFP reporter library (Figure 3)

J3 upstream region (without 26 bases at 3'), N26, BBa_J23117 promoter, Bujard RBS, sfGFP, BBa_B0015 terminator. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAANNNNNNNNNNNNNNNNNNNNNNNNTTGACAGCTAGCTCAGTCCTAGGGATTGTGCTAGCGAATT
CATTAAAGAGGAGAAAGGTACCATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTT
GAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACG
GAAAACTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTAC
TCTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGT
GCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTG
CTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGA
TGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAGACAAA
CAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAG
ACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGAC
ACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCT
GCTGGGATTACACATGGCATGGATGAGCTCTACAAATAAGGATCCAAACTCGAGTAAGGATCTCCAGGCA
TCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCT
CTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA

Modified synthetic promoters

Promoters with different basal expression levels (Figure 3)

Annotations:

J3 upstream region, variable promoter (BBa_J231xx), Bujard RBS, Start codon of mRFP. The J106 target site is underlined

BBa_J23105

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC

ACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtactatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23106

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtatagtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23107

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagccctaggtattatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23108

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGActgacagctagctcagtcctaggtataatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23109

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttacagctagctcagtcctagggactgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23110

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttacggctagctcagtcctaggtacaatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23111

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAttgacggctagctcagtcctaggtatagtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23112

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGActgatagctagctcagtcctagggattatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23113

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGActgatggctagctcagtcctagggattatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23114

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttatggctagctcagtcctaggtacaatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23115

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAtttatagctagctcagcccttggtacaatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23118

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAttgacggctagctcagtcctaggtattgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

BBa_J23119

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctaggtataatgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

Promoters regulated by alternative sigma factors (Figure 3)

Annotations:

J1 upstream region, variable promoter (color coded), Bujard RBS, Start codon of mRFP. The J106 target site is underlined.

J1-*sodCp* promoter

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTG
ACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCG
TCCTTTGGGTTCCACCGGATACCTCCGGCCGTTcaaaaatgtgtcacTGGTTTACACTtattcagGGAAT
TCATTAAAGAGGAGAAAGGTACCATG

In the J1-*sodCp* promoter, we deleted 3 bp between the J1 sequence and the *sodCp* minimal promoter to maintain the same -80 bp spacing between the target site and the J109 target site as in the rpoD promoter. Without this deletion, activation from the *sodCp* promoter is substantially weaker (not shown), likely due to the sensitive positioning requirements described in Figure 4.

J1-*glnAp2* promoter

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTG
ACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCG
TCCTTTGGGTTCCACCGGATACCTCCGGACTGGCACagatttCGCTTtatctttttTacggcgacGAATT
CATTAAAGAGGAGAAAGGTACCATG

J1-*rdgBp* promoter

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTG
ACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCG
TCCTTTGGGTTCCACCGGATACCTCCGGACTTGTAAaaggcgaacttggcCGCCACaaacaaattGAATT
CATTAAAGAGGAGAAAGGTACCATG

J1-*yieEp* promoter

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTG
ACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCG
TCCTTTGGGTTCCACCGGATACCTCCGGACGAACTTttagccgctttagtctgtcCATCAttccAGAATT
CATTAAAGAGGAGAAAGGTACCATG

J3(tetO)-J23117_mRFP1 reporter (Figure 3)

J3 upstream region (without 19 bases at 3'), tetO, BBa_J23117 promoter, Bujard RBS, ATG of mRFP1. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCACTCTATCGTTGATAGAGTttgacagctagctcagtcctagggattgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

J1-J23117 promoter with shifted bases (Figure 4, Supplementary Figure 5 and subsequent)

For promoters shifted by fewer than 12bp, bases were removed from the 5' end of the inserted sequence (starting with C).

J1 upstream region, inserted sequence (12 bases), BBa_J23117 promoter, Bujard RBS, ATG of mRFP1. The J106 target site is underlined.

GCCTACGGTATCCACCGGAGACCTATGGCAGCCTCCGGCCGCCATAGGACACCTTTGGTTGCCAAGGGTG
ACCTATGGTGACCATGGGCCACCACGGGCGACCTCAGGTATCCTGCGGTGTCCTGCGGTTACCAAAGGCG
TCCTTTGGGTTCCACCGGATACCTCCGGACCTAGATGCGCTCttgacagctagctcagtcctagggattg
tgctagcGAATTCATTAAAGAGGAGAAAGGTACCATG

J3-J23117 promoter with shifted bases (Supplementary Figure 5 and subsequent)

For promoters shifted by fewer than 12bp, bases were removed from the 5' end of the inserted sequence (starting with T). For promoters with negative shifts, no sequence was inserted between the J3 upstream region and BBa_J23117, and bases were removed from the 3' end of the J3 upstream region (starting with A).

J3 upstream region, inserted sequence (12 bases), BBa_J23117 promoter, Bujard RBS, ATG of mRFP1. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGATGAGATGCGCTCttgacagctagctcagtcctagggattg
tgctagcGAATTCATTAAAGAGGAGAAAGGTACCATG

Modified J3-J23117 promoter with 6 adjacent PAMs around -81 bp to TSS (Supplementary Figure 5)

J3 upstream region, BBa_J23117 promoter, Bujard RBS, Start codon of mRFP. Region with additional PAM sites inserted. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACCCCCCCTACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

Modified J3-J23117 promoter with non-NGG PAMs around -81 bp to TSS (Figure 5A)

J3 upstream region, BBa_J23117 promoter, Bujard RBS, Start codon of mRFP, region with modified PAMs. The J306 target site is underlined.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTNNNACGGAGCGTTCTGGAC
ACAACGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

Promoter to demonstrate dxCas9(3.7) versatility (Figure 5B)

J3 upstream region, BBa_J23117 promoter, Bujard RBS, Start codon of mRFP. The modified target site (M) is underlined. AGT PAM site.

AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCGATGTCATAATCATCGCTACGAGCT
GTGAAAGATGCATAAAGCTCGTACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACACTGCTGACACTTC
TGCTCGTCGTCTTGAAGTTGCGATTATAGAttgacagctagctcagtcctagggattgtgctagcGAATT
CATTAAAGAGGAGAAAGGTACCATG

Supplementary references

1. Keseler, I. M. *et al.* The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res* **45**, D543–D550 (2017).
2. Gama-Castro, S. *et al.* RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res* **44**, D133–D143 (2016).
3. Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nat Commun* **9**, 2489 (2018).
4. Griffith, K. L. & Wolf, R. E. A comprehensive alanine scanning mutagenesis of the *Escherichia coli* transcriptional activator SoxS: Identifying amino acids important for DNA binding and transcription activation. *Journal of Molecular Biology* **322**, 237–257 (2002).
5. Ryoichi Arai, T. N., Hiroshi Ueda, Atsushi Kitayama, Noriho Kamiya. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering, Design and Selection* **14**, 529–532 (2001).
6. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).
7. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
8. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **41**, 7429–7437 (2013).
9. Johnny H. Hu, D. R. L., Shannon M. Miller, Maarten H. Geurts, Weixin Tang, Liwei Chen, Ning Sun, Christina M. Zeina, Xue Gao, Holly A. Rees, Zhi Lin. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
10. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods* **3**, 623–628 (2006).
11. Donghyuk Kim, B. Ø. P., Jay Sung-Joong Hong, Yu Qiu, Harish Nagarajan, Joo-Hyun Seo, Byung-Kwan Cho, Shih-Feng Tsai. Comparative analysis of regulatory elements between *Escherichia*

coli and *Klebsiella pneumoniae* by genome-wide transcription start site profiling. *PLoS Genetics* **8**, e1002867 (2012).